THREE ESSAYS ON SEMIPARAMETRIC ECONOMETRICS: THEORY AND APPLICATION

A Dissertation

by

HONGJUN LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Co-Chair of Committee,	Jianhua Huang
Committee Members,	Dennis W. Jansen
	Ke-Li Xu
Head of Department,	Timothy Gronberg

May 2014

Major Subject: Economics

ABSTRACT

This dissertation aims at investigating the theory and application of semiparametric econometrics. I first inspect the selection of optimal bandwidth using the cross-validation method for the kernel estimation of cumulative distribution/survivor functions. Then, I analyze the determination of the number of factors with the methods of principal component and information criteria. I also show the application of semiparametric methods to "purchasing power parity" puzzle.

Firstly, I propose a data-driven least squares cross-validation method to optimally select smoothing parameters for the nonparametric estimation of cumulative distribution/survivor functions. The general multivariate covariates can be continuous, discrete/ordered categorical or a mix of either. I establish the asymptotic optimality of least squares cross-validation method. Also, I show that the estimators of cumulative distribution/survivor functions using the smoothing parameters selected by the proposed method is asymptotically normally distributed. Monte Carlo simulation verifies the finite-sample properties of the least squares cross-validation method.

Secondly, I provide some discussions on the econometric theory for factor models of large dimensions where the number of factors (r) is allowed to increase as the two dimensions, cross-sections (N) and time dimensions (T) increase. I mainly focus on the determination of the number of factors. I extend the existing panel criteria to high dimension case where r may be increasing with N or T. I show that the number of factors can be consistently estimated using the criteria. Also, Monte-Carlo simulation demonstrates the finite sample properties of the proposed estimating method.

Lastly, I consider an empirical application of semiparametric econometrics to the problem of purchasing power parity (hereafter PPP) hypothesis test. Traditional linear cointegration tests of PPP hypothesis often lead to rejection of the PPP hypothesis. More recent studies allowing for some sort of nonlinearity in econometric modelings suggest mixed results and leave this problem as an unresolved issue. Therefore, I analyze PPP hypothesis within a semiparametric framework using the varying coefficient model with integrated variables, which can capture the nonlinearity of the economic structures. Applying the semiparametric functional cointegration test method, I conduct the cointegration test of PPP hypothesis between U.S. and Canada, U.S. and Japan, and U.S. and U.K., respectively to test the PPP hypothesis. In contrast to the usual findings based on linear model PPP hypothesis testing, the semiparametric model based tests provide supporting evidence of the PPP hypothesis.

DEDICATION

to my family, my teachers and my friends.

ACKNOWLEDGEMENTS

First of all, I would like to thank my committee chair, Dr. Qi Li, committe co-chair, Dr. Jianhua Huang, and my committee members, Dr. Dennis Jansen and Dr. Ke-Li Xu for their guidance and support throughout the course of this research. Dr. Qi Li introduced me to the wonderful field of theoretical econometrics. We met almost every week and discussed papers and ideas at his office during the past two years. He also provided numerous evaluations and valuable feedback of my research to help me improve the final product. Dr. Jianhua Huang, Dr. Dennis Jansen and Dr. Ke-Li Xu shared their ideas with me, taught me many important research skills and offered constructive comments to my work. What I have learnt from them will be my fortune for my whole life.

I also want to extend my gratitude to other people who helped me a lot during my study of economics. Dr. Hae-shin Hwang shared with me his precious life experience and offered me great encouragement. Dr. Ryo Jinnai and Dr. Yuzhe Zhang showed me the research of macroeconomics step by step. Dr. Li Gan did his best to help me with both my study and life. Rodrigo Velez also provided helpful guidance and support. Dr. Da Lu of Tsinghua University and Dr. Minsoo Lee of Asian Development Bank (ADP) instructed me on my study when I was in China.

Thanks also go to my friends and colleagues and the other department faculty and staff for making my time at Texas A&M University a great experience. Finally, thanks to my parents for their encouragement and to my girlfriend, Miss Lu Fu for her patience, support and love.

NOMENCLATURE

CDF	Cumulative Distribution Function
CV	Cross-Validation
DGP	Data Generating Process
LS-CV	Least Squares Cross-Validation
PDF	Probability Density Function
PPP	Purchasing Power Parity

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Model specification is a critical issue for economic analysis, as a misspecified model may contaminate the analytical results dramatically. If one specifies the model as linear while the relationship between the dependent variable and explanatory variables is actually nonlinear, the estimation and inference reasults can be misleading. Semiparametric/nonparametric methods provide one way to solve the misspecification problem by estimating and modeling the underlying structures of the economic relationship simultaneously .

This dissertation tries to investigate the theory and application of semiparametric econometrics. The first discussion concerns the problem of bandwidth selection. I propose a data-driven least squares cross-validation method to select the smoothing parameters for the nonparametric estimation of cumulative distribution/survivor functions optimally. I put no restrictions on the data-type of variables. The general multivariate covariates can be continuous, discrete/ordered categorical or a mix of either. I establish the asymptotic optimality of least squares cross-validation method. Also, I show that the estimators of cumulative distribution/survivor functions is asymptotically normally distributed if one selects the smoothing parameters through the proposed method. The Monte Carlo simulation verifies the finite-sample properties of the least squares cross-validation method.

I also inspect the econometric theory for determining the number of factors where the factor models are of large dimensions. I allow the number of factors (r) to increase as the two dimensions, cross-sections (N) and time dimensions (T) increase. This is a typical situation one may come across while conducting financial or macroeconomic analyses. I mainly focus on the determination of the number of factors by extending the criteria suggested in Bai and Ng (2002) to high dimension case where r may be increasing with N or T. I show that the number of factors can be consistently estimated using the proposed criteria. Also, Monte-Carlo simulation demonstrates the finite sample properties of the proposed estimating method and suggests that the panel criteria perform well if the sample sizes are not too small, i.e., $\min\{N, T\} \geq 60$.

In the application part, I show an empirical application of semiparametric econometrics to the problem of purchasing power parity (hereafter PPP) hypothesis test. Traditional linear cointegration tests of PPP hypothesis often lead to rejection of the PPP hypothesis as the relationship between exchange rate and price levels may be nonlinear. More recent studies allowing for some sort of nonlinearity in econometric modelings (e.g., Michael et al. 1997) suggest mixed results and leave this problem as an unresolved issue. In this section, I investigate PPP hypothesis within a semiparametric framework using the varying coefficient model with integrated variables as considered by Cai, Li, and Park (2009), and Xiao (2009). Applying the cointegration test suggested by Xiao (2009), I conduct the cointegration test of PPP hypothesis between U.S. and Canada, U.S. and Japan, and U.S. and U.K., respectively to test the PPP hypothesis. In contrast to the usual findings based on linear model PPP hypothesis testing, the semiparametric model based tests provide supporting evidence of the PPP hypothesis.

The rest of this dissertation is organized as follows. Section 2 analyzes the smoothing parameters selection problem using least squares cross-validation methods. Section 3 studies the method to determine the number of factors when it may increase with sample sizes. Section 4 shows the application of varying coefficient method to the test of "purchasing power parity" hypothesis. The last section summarizes this dissertation and draws some conclusions. All the proofs of the theorems for each section are given in the appendixes.

2. CROSS-VALIDATED MIXED DATATYPE BANDWIDTH SELECTION FOR NONPARAMETRIC CUMULATIVE DISTRIBUTION/SURVIVOR FUNCTIONS

We propose a data-driven least squares cross-validation method to optimally select smoothing parameters for the nonparametric estimation of cumulative distribution/survivor functions. We allow for general multivariate covariates that can be continuous, discrete/ordered categorical or a mix of either. We provide asymptotic analysis, examine finite-sample properties via Monte Carlo simulation, and consider an illustration involving nonparametric copula modeling.

2.1 Introduction

Though the kernel estimation of multivariate probability density functions (PDFs) has received much attention in the literature, the estimation of multivariate cumulative distribution functions (CDFs) has received less attention (see Liu and Yang 2008 and Li and Racine 2008). Furthermore, data-driven methods for bandwidth selection are noticeably absent from the practically important setting involving a mix of discrete and continuous variables. This paper aims to fill this gap with a simple and practical cross-validation method for multivariate bandwidth selection for unconditional mixed data CDFs. See also related work by Bashtannyk and Hyndman (2001), Bowman, Hall and Prvan (1998) among others.

One important aspect of this approach is that, unlike its PDF-based counterpart, the multivariate nonparametric CDF estimator has a $(n^{-1/2} \log n)$ rate of convergence which does not depend on the number of covariates thus does not suffer from the curse of dimensionality that plagues many related nonparametric estimators. This renders it ideally suited for a range of scenarios, from the nonparametric estimation of copula functions to novel tests for multivariate data (i.e. smooth Cramer-von Mises/Kolmogorov-Smirnov types of tests).

2.2 Multivariate CDF Bandwidth Selection

We consider the case for which y is a vector containing mixed discrete/ordered categorical and continuous variables, and presume interest lies in smooth estimation of $F(y)$, the cumulative distribution function or $S(y) = 1 - F(y)$, the survivor function (results below transfer immediately to the survivor function given that it is a trivial transformation of the cumulative distribution function). Let $y = (y^c, y^d)$, where y^c is a q-dimensional continuous random vector, and where y^d is an r-dimensional discrete random vector. Let y^d_{is} (y^d_s) denote the s^{th} component of y^d_i (y^d), $s = 1, \ldots, r$; $i = 1, \ldots, n$, where n is the sample size. We assume that y_s takes values in $\{0, 1, \ldots, c_s - 1\}$, where $c_s \geq 2$ is a positive integer. Let λ denote the bandwidth for a discrete variable. For the discrete variables we use the kernel $l(Y^d_{is}, y^d_s, \lambda_s) = \lambda_s^{|Y^d_{is} - y^d_s|}$ (with $\lambda^0_s = 1$ and $0^0 = 1$), where $\mathbf{1}(A) = 1$ if A holds, and 0 otherwise. I write the product (discrete variable) kernel as $L_\lambda(y^d_i, y^d, \lambda) = \prod_{s=1}^r l(y^d_{is}, y^d_s, \lambda_s)$. The product kernel function used for the continuous variables is given by $W_h(y^c_i, y^c) = \prod_{s=1}^q h_s^{-1} w((y^c_{is} - y^c_s)/h_s)$, where $w(\cdot)$ is a univariate kernel function for a continuous variable. y^c_{is} (y^c_s) denotes the s^{th} component of y^c_i (y^c) and h_s is the bandwidth associated with y^c_s. The kernel function for the vector of mixed variables $y = (y^c, y^d)$ is simply the product of $W_h(\cdot)$ and $L_\lambda(\cdot)$ given by $K_\gamma(y_i, y) = W_h(y^c_i, y^c) \times L_\lambda(y^d_i, y^d, \lambda)$, where $\gamma = (h, \lambda)$.

We use $F(y)$ to denote the unconditional CDF of a multivariate Y which may contain a mix of discrete and continuous covariates, and consider a kernel-based estimator defined by

$$\hat{F}(y) = n^{-1} \sum_{j=1}^{n} G_\gamma(y, y_j), \tag{2.1}$$

where $G_\gamma(y, y_j)$ is a multivariate mixed data cumulative distribution kernel function obtained from $K_\gamma(y_i, y)$ via $G_\gamma(y, y_j) = \int K_\gamma(y_i, y)\, dy$ where $\int \dots dy$ is taken to mean $\sum_{y_d} \int \dots dy^c$. Theoretical properties of this estimator when y is continuous only can be found in Liu and Yang (2008), where they shoId that the estimator follows the same pointwise asymptotically normal distribution as the empirical CDF, while the smooth kernel estimator has asymptotically smaller mean integrated squared error than the empirical CDF, and converges to the true CDF uniformly almost surely at a rate of $(n^{-1/2} \log n)$. As we demonstrate for the mixed-data multivariate case, a dimension-free rate also holds rendering this estimator ideally suited to a range of potential applications. The missing component is, naturally, a data-driven bandwidth selector (though Liu and Yang (2008) propose a plug-in type bandwidth selector for the continuous only case, it is not applicable here).

Following Bowman, Hall and Prvan (1998), we could choose bandwidths by minimizing the cross-validation function

$$CV(\gamma) = n^{-1} \sum_{i=1}^{n} \int \left\{ \mathbf{I}(y_i \le y) - \hat{F}_{-i}(y) \right\}^2 dy. \tag{2.2}$$

The univariate continuous y version of this statistic was proposed and studied by Bowman, Hall and Prvan (1998). We could generalize this to the multivariate mixed data setting (if an element of y is discrete, then one could replace $\int dy$ by $\sum_{y \in D_y}$ in (2.2)). One drawback with such an approach is purely practical – numerical integration over the continuous variables would be required, and when there exist more than one continuous variable the reliance on multivariate numerical integration would present a barrier for adoption.

Rather than generalizing (2.2) to the multivariate mixed data case, we instead precede with the cross-validation function

$$CV(h) = \frac{1}{nn_j} \sum_{i=1}^{n} \sum_{j=1}^{n_j} \left\{ \mathbf{I}(y_i \le y_j^e) - \hat{F}_{-i}(y_j^e) \right\}^2, \tag{2.3}$$

where y_j^e, $j = 1, \ldots, n_j$, denotes evaluation points. The number of evaluation points could be fixed at, say, $n_j = 100$. This grid of evaluation points plays a role not unlike the number/position of points used for numerical integration of (2.2). This statistic is appealing from a practical standpoint as it would scale Ill with respect to n and sidesteps the need for multivariate numerical integration. For the y^e we advise using an equi-quantile grid in each dimension derived from the empirical distribution of the marginals of y (i.e. the set of quantiles of each variable in y for $\tau = 0, 1/n_j, 2/n_j, \ldots, 1$). Note that the dependence structure is embodied in the estimation data (i.e. the y_is), so evaluating on an equi-quantile grid in no way restricts the underlying structure nor does it appear to have any deleterious effect on the resulting bandwidths (relative to that where $y^e = y$ and $n_j = n$).

Simulation studies similar to those undertaken in Li, Lin and Racine (2013) indicate that there is no perceptible loss arising from using the summation versus integral variant of a similar (univariate y, conditional) statistic. We pursue (2.3) for this and the reasons outlined above.

2.3 Theoretical Properties

In this section, we analyze the theoretical properties of the cross-validation bandwidths selection methods. We make the following assumptions for our analyses. All the proofs are saved in Appendix A.

Condition 1. $\{Y_j^c, Y_j^d\}_{j=1}^n$ *are independent and identically distributed as* (Y^c, Y^d), $F(y^c|y^d)$ *have uniformly continuous third order partial derivative functions with respect to* y^c.

Condition 2. *$w(\cdot)$ is a non-negative, symmetric and bounded second order kernel function with $\int_{-\infty}^{\infty} w(v)v^2 dv$ and $\int_{-\infty}^{\infty} G(v)w(v)vdv$ being finite constants, where $G(v) \equiv \int_{-\infty}^{v} w(u)du$.*

Condition 3. *As $n \to \infty$, $h_s \to 0$ for $s = 0, 1, \ldots, q$, $\lambda_s \to 0$ for $s = 1, \ldots, r$.*

For $s = 1, \ldots, q$, $F_s(y^c, y^d) = \partial F(y^c, y^d)/\partial y_s^c$, $F_{ss}(y^c, y^d) = \partial^2 F(y^c, y^d)/\partial y_s^{c2}$, $\kappa_2 = \int w(v)v^2 dv$, $\alpha_0 = 2\int vG(v)w(v)dv$. The next lemma gives the leading terms for the estimation MSE of $\hat{F}(y)$.

Lemma 2.3.1 *Under conditions 1-3, we have*

$$\begin{aligned}\text{MSE}\left(\hat{F}(y)\right) =& -\alpha_0 \sum_{s=1}^{q} \frac{h_s}{n} F_s(y^c, y^d) + 2\sum_{s=1}^{r} \frac{\lambda_s}{n}\left[B_{2s}(y^c, y^d) - F(y)B_{1s}(y^c, y^d)\right] \\ &+ \left(\frac{\kappa_2}{2}\sum_{s=1}^{q} h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s B_{1s}(y^c, y^d)\right)^2 + \frac{1}{n}F(y)(1 - F(y)) \\ &+ o\left(\sum_{s=1}^{r}\left(\frac{\lambda_s}{n} + \lambda_s^2\right) + \sum_{s=1}^{q}\left(\frac{h_s}{n} + h_s^4\right)\right) \qquad (2.4)\end{aligned}$$

where $B_{1s}(y^c, y^d) \equiv E_{y_j^d}\left\{\sum_{z^d \le y^d} \mathbf{1}(y_{-sj}^d = z_{-s}^d) d(y_{sj}^d, z_s^d) F(y^c|y_j^d)\right\}$ with $\mathbf{1}(A)$ as the indicator function, and $B_{2s}(y^c, y^d) \equiv E_{y_j^d}\left\{\sum_{z_d \le y^d} \mathbf{1}(y_j^d = z^d) d(y_{sj}^d, z_s^d) F(y^c|y_j^d)\right\}$, for $s = 1, \ldots, r$.

Denote the leading terms of MSE $\left(\hat{F}(y)\right)$ as $\text{MSE}_L\left(\hat{F}(y)\right)$, which is

$$\begin{aligned}\text{MSE}_L\left(\hat{F}(y)\right) =& -\alpha_0 \sum_{s=1}^{q} \frac{h_s}{n} F_s(y^c, y^d) + 2\sum_{s=1}^{r} \frac{\lambda_s}{n}\left[B_{2s}(y^c, y^d) - F(y)B_{1s}(y^c, y^d)\right] \\ &+ \left(\frac{\kappa_2}{2}\sum_{s=1}^{q} h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s B_{1s}(y^c, y^d)\right)^2 + \frac{1}{n}F(y)(1 - F(y))\end{aligned}$$

Next, we present the result concerning the relation betIen the leading terms of $CV(\cdot)$ and the leading terms of MSE $\left(\hat{F}(y)\right)$.

Theorem 2.3.1 *Assuming that conditions 1 to 3 hold, then the leading term of* $CV(\cdot)$ *is given by* $CV_L(\cdot)$ *which is defined as follows (where* $\int dy = \sum_{y^d \in D_y} \int dy^c$*)*

$$CV_L(\gamma) = \int \mathrm{MSE}_L\left(\hat{F}(y)\right) dy \tag{2.5}$$

As the data type of y is not restricted for our proof, this theorem holds for any type of y satisfying conditions 1-3, regardless of whether y is discrete, continuous, or mixed. Also, this theorem suggests that the CV selected bandwidth is asymptotically optimal as the leading term from CV function is equal to MSE of $\hat{F}(y)$. Li, Lin and Racine (2013) obtain a similar result for conditional CDF estimation.

Let $\hat{h}_s$ for $s = 1, \ldots, q$, and $\hat{\lambda}_s$ for $s = 1, \ldots, r$ denote the bandwidths that minimize $CV(\gamma)$. Denote $h_s^0 = a_s^0 n^{-1/3}$ and $\lambda_s^0 = b_s^0 n^{-2/3}$ as the bandwidths that minimize the leading terms of the Iighted integrated estimation MSE. With the similar discussion as Li, Lin and Racine (2013), we can show the convergence rate of the CV selected bandwidths by combining the results of Lemma 2.3.1 and Theorem 2.3.1 as follows:

Theorem 2.3.2 *Under conditions 1-3, we have*

1. $\frac{\hat{h}_s}{h_s^0} \xrightarrow{p} 1$, *for* $s = 1, \ldots, q$;
2. $\frac{\hat{\lambda}_s}{\lambda_s^0} \xrightarrow{p} 1$, *for* $s = 1, \ldots, r$;

where $a_s^0 (s = 1, \ldots, q)$ *are positive constants, and* $b_s^0 (s = 1, \ldots, r)$ *are non-negative constants.*

Given the asymptotic behavior of the smoothing parameters, we can discuss the asymptotic distribution of the estimator $\hat{F}(y)$, which is given in the next theorem.

Theorem 2.3.3 (Asymptotic Distribution) *Under conditions 1-3, we have*

$$\sqrt{n}\left[\hat{F}(y) - F(y)\right] \xrightarrow{d} N(0, V) \tag{2.6}$$

where $V = F(y)(1 - F(y))$.

Theorem 2.3.3 shows that the estimator $\hat{F}(y)$ using the CV selected bandwidths is asymptotically normally distributed, and the convergence rate is $\sqrt{n}$.

2.4 Monte Carlo Simulation

We simulate data from a multivariate normal distribution of dimension k with mean vector $\mu = (0, \ldots, 0)'$ and covariance matrix Σ a matrix with ones on the diagonal and $\rho = 0.25$ on the off-diagonals. We compare the multivariate PDF least-squares approach of Li and Racine (2003). We set $n_j = 100$, and draw $M = 1,000$ replications from this DGP and consider four estimators, namely, the proposed approach, that optimal for unconditional PDFs (Li and Racine, 2003), the nonsmooth multivariate empirical distribution function (which is (2.1) with $\gamma = 0$), and the Oracle estimator (i.e. that based on information of the true DGP with the mean vector and covariance matrix replaced by their sample counterparts).

We report the efficiency of the PDF-based, empirical, and Oracle estimators as measured by their median MSE over all M replications relative to that of the proposed estimator in Table 2.1, where numbers greater than one indicate inferior MSE performance of the estimator named at the top of each column. We vary n from 100 through $1,600$ and k from 1 through 5. Naturally, the Oracle estimator will dominate as this makes use of information regarding the underlying DGP. We expect the performance of the PDF-based estimator to fall as k increases while that of the proposed and empirical CDF estimators ought not (this can be ascertained by the relative efficiency of the Oracle estimator).

Table 2.1: Relative Efficiency of the Proposed CDF Bandwidth Approach

k	n	h_{pdf}	$h=0$	Oracle
1	100	1.16	1.22	0.60
1	200	1.16	1.19	0.63
1	400	1.26	1.19	0.58
1	800	1.31	1.16	0.60
1	1600	1.31	1.12	0.59
2	100	1.13	1.26	0.62
2	200	1.27	1.25	0.57
2	400	1.37	1.20	0.59
2	800	1.47	1.18	0.56
2	1600	1.53	1.16	0.55
3	100	1.17	1.29	0.57
3	200	1.43	1.27	0.57
3	400	1.62	1.25	0.55
3	800	1.90	1.17	0.56
3	1600	2.04	1.13	0.50
4	100	1.36	1.26	0.54
4	200	1.78	1.30	0.56
4	400	1.99	1.22	0.49
4	800	2.60	1.16	0.46
4	1600	3.37	1.17	0.44
5	100	1.54	1.25	0.51
5	200	2.13	1.30	0.53
5	400	2.71	1.14	0.52
5	800	4.10	1.18	0.44
5	1600	6.01	1.13	0.50

2.5 Application to Nonparametric Copula Models

Nonparametric estimation of copula has been addressed in Chen and Huang (2007) (We direct the reader to Nelsen, 2006 for an authoritative treatment of copula). Nonparametric estimation of copula involves estimation of a joint CDF, hence bandwidth selection becomes of paramount importance. Given a bivariate distribution function H defined over two random variables X and Y with continuous marginals F and G, I consider estimation of a bivariate nonparametric copula $C(u,v) = H(F^{-1}(u), G^{-1}(v))$ where the data are generated a joint normal with $\rho_{xy} = 0$ and $\rho_{xy} = 0.99$. We consider one draw of size $n = 1,000$ and estimate the bandwidths using the approach proposed in this paper (estimated bandwidths Ire $h_x = 0.1670367$ and $h_y = 0.1502703$ for $\rho_{xy} = 0$, $h_x = 0.04028061$ and $h_y = 0.03499221$ for $\rho_{xy} = 0.99$, $h_x = 0.07739613$ and $h_y = 0.03985951$ for $\rho_{xy} = -0.99$, and a second order Gaussian kernel was employed).

We plot the resulting contour plot (along with the true DGP), scatter plot, and perspective plot in Figures 2.1 through 2.3.

Figures 2.1-2.3 suggest that the smooth nonparametric copula estimate is capable of delivering a faithful representation of the unknown copula.

2.6 Concluding Remarks

We consider a cross-validated approach towards computation of bandwidths for kernel estimation of CDF functions that admits both discrete and continuous variables. Theoretical underpinnings are provided, while an application to estimation of a bivariate copula demonstrates the potential utility of the proposed method for practitioners. The approach has the added benefit of being computationally appealing due to the manner in which the cross-validation function can be expressed. An implementation in R programming language is available in the np package (See

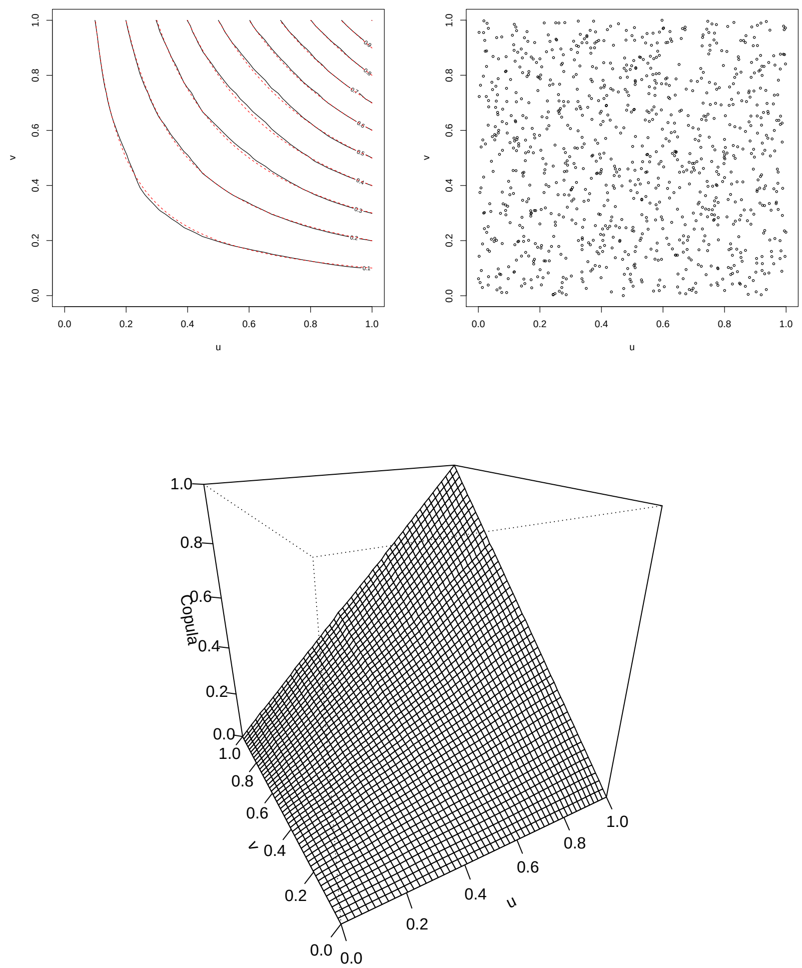

Figure 2.1: Bivariate Copula Plots: $\rho_{xy} = 0$, $n = 1,000$

Hayfield and Racine, 2008).

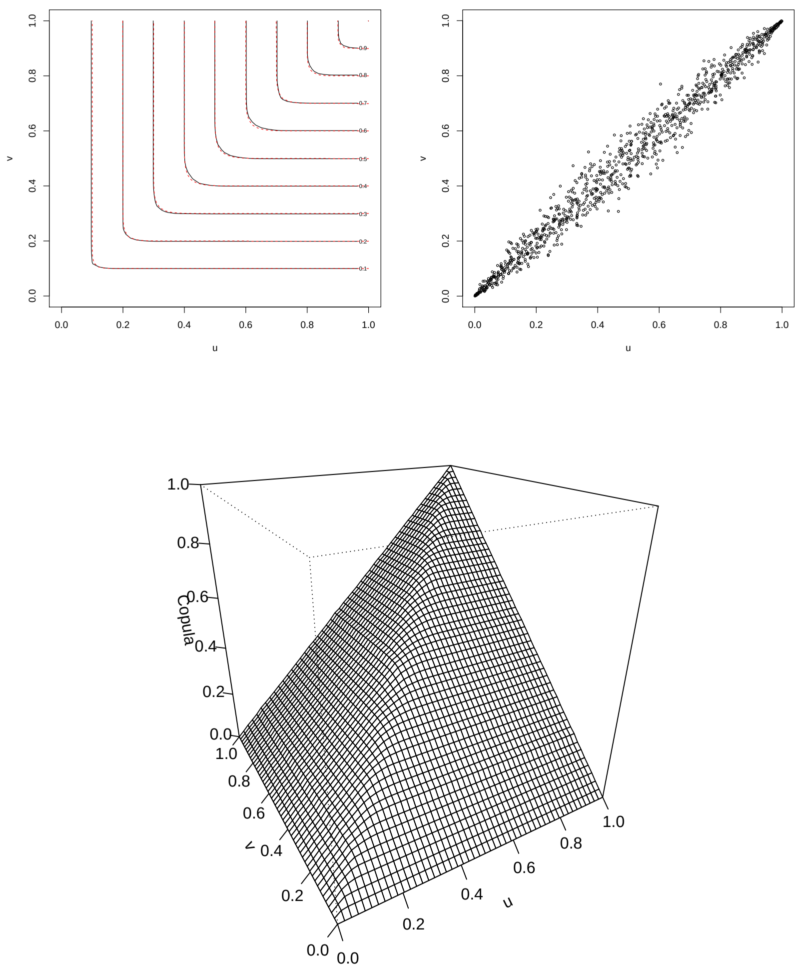

Figure 2.2: Bivariate Copula Plots: $\rho_{xy} = 0.99$, $n = 1,000$

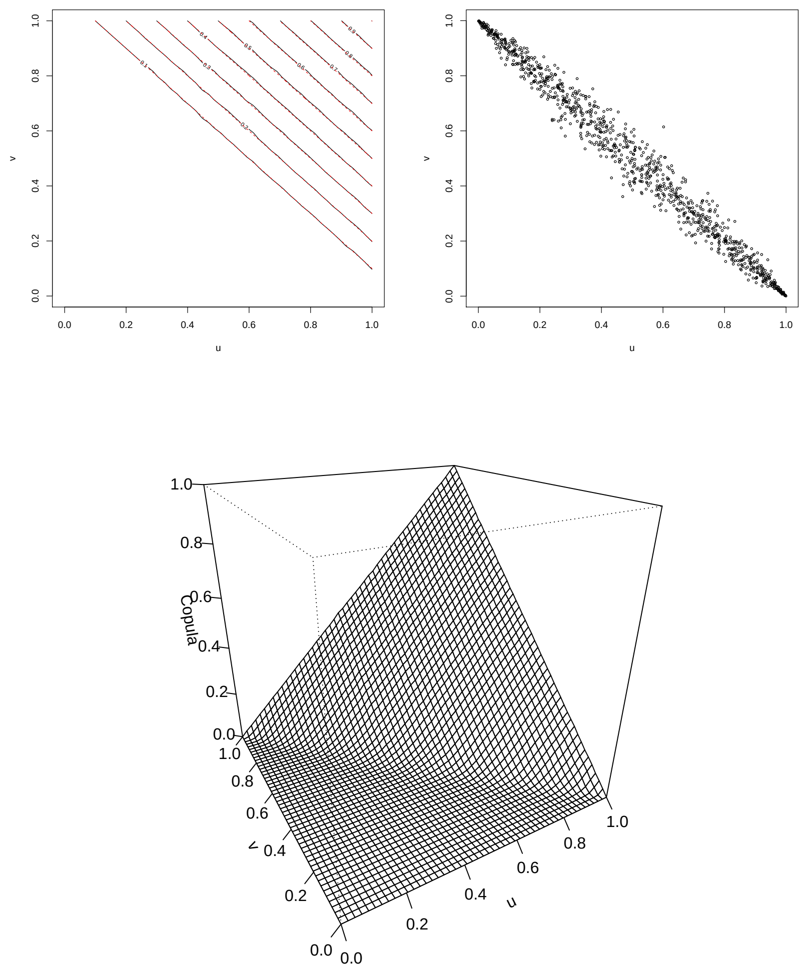

Figure 2.3: Bivariate Copula Plots: $\rho_{xy} = -0.99$, $n = 1,000$

3. DETERMINING THE NUMBER OF FACTORS WHEN THE NUMBER OF FACTORS CAN INCREASE WITH SAMPLE SIZES

Correctly specifying the number of factors (r) is a fundamental issue for the application of factor models. In this section we develop an econometric method to estimate the number of factors in factor models of large dimensions where the number of factors is allowed to increase as the two dimensions, cross-section size (N) and time period (T) increase. Using similar information criterion as proposed by Bai and Ng (2002), we show that the number of factors can be consistently estimated using the criteria. We propose a new procedure that avoids over estimating the number of factors while allowing for one to search for possible number of factors over a wide range of positive integers so that it also avoids underestimation of the number of factors. We conduct Monte-Carlo simulation to investigate the finite sample properties of the proposed approach.

3.1 Introduction

Factor models have been widely used in economic analyses such as forecasting economic variables, estimating variance-covariance matrix with high dimension data, and estimating average treatment effects. In practice a few common factors may capture the variations of a large number of economic variables. In the finance literature, the arbitrage pricing theory (APT) of Ross (1976) assumes that a small number of factors can be used to explain a large number of asset returns. Stock and Watson (1998, 1999) consider forecasting inflation with diffusion indices ("factors") constructed from a large number of macroeconomic series. Gregory and Head (1999) and Forni, Hallin, Lippi, and Reichlin (2000) find that cross country variations have common components. Fan, Liao and Mincheva (2011), and Fan, Liao and Mincheva

(2013) use factors model to estimate high dimensional variance-covariance matrix. Factor models can be used to evaluate the impacts of various polices. By assuming that the cross-sectional correlations for all the units are attributed to the presence of some (unobserved) common factors, Hsiao, Ching and Wan (2012) offer a panel data method to construct the counterfactuals and to measure average treatment effects of policy interventions based on factor models.

A fundamental issue of factor models is the correct specification of the number of factors, r. When the number of factors is fixed, Bai and Ng (2002), Onatski (2009), Anh and Horenstein (2013), among others, have developed various approaches to consistently estimate the number of factors. But many empirical findings suggest that the number of factors may increase as the dimensions of the data N increases, or T increases. For many empirical analyses, the estimated number of factors ranges from one to more than ten, see Ludvigson and Ng (2009), Giannone, Reichlin and Sala (2005) and Forni and Gambetti (2010). This suggests that the number of factors may dependent on sample sizes. One reason that the number of factors may increase with sample sizes is structure break, like new factors may emerge after economic environments change. Using Bai and Ng's (2002) information criteria, Ludvigson and Ng (2007) find that the factor structure of their financial dataset comprising of 172 ($N = 172$) series quarterly financial indicators spanning the first quarter of 1960 through the fourth quarter of 2002 ($T = 172$) can be well described by 8 ($r = 8$) common factors. Jurado, Ludvigson and Ng (2013) update monthly version of the 147 financial time series used in Ludvigson and Ng (2007), and combine them with an updated version of 132 monthly macroeconomic series used in Ludvigson and Ng (2010). They find that 12 ($r = 12$) common factors can capture the variations of this new dataset with 279 series ($N = 279$) spanning the period 1959:01-2011;12 ($T = 636$). Hence, Jurado, Ludvigson and Ng's (2013) finding supports the argument that the number of factors may increase as sample increases.

Assuming that the number of factors r is fixed, there are many papers in the literature analyzing the problem of determining the number of factors. Some of them not only fix the number of factors, but also impose restrictions the dimensions N and T, such as Lewbel (1991), Donald (1997), Cragg and Donald (1997), Connor and Korajczyk (1993), Forni and Reichlin (1998) and Stock and Watson (1998). Imposing no restriction on the relation between N and T except that both N and T are assumed to be large, Bai and Ng (2002) treat the determination of the number of factors as a model selection problem, they propose some criteria and show that the number of factors can be consistently estimated by minimizing the proposed criteria. Onatski (2009) develops a test of the null of k_0 factors against the alternative that the number of factors r is $k_0 < r \leq k_1$ for some finite positive integer k_1. Onatski also describes the asymptotic distribution of the test statistic with critical values tabulated. Onatski (2010) suggests to determine the number of factors from empirical distribution of eigenvalues of sample covariance matrix. Ahn and Horenstein (2013) exploit the fact that the r largest eigenvalues of the variance matrix of N response variables grow unboundedly as N increases, while the other eigenvalues remain bounded to estimate the number of factors. The main difference between our analyses and the existing work is that we consider the problem of determining the number of factors in a factor model where the number of factors is allowed to increase as N or T increases.

Specifically, this section is designed to provide an approach which enables one to estimate the number of factors consistently when the number of factors is allowed to increase as $N, T \to \infty$. We extend the method of Bai and Ng (2002) to penalize the number of factors with a penalty function which is determined by the sample sizes, N and T, as well as the maximum possible number of factors allowed in the estimation. As the factors are unobserved, the estimating procedure takes two steps. First, assuming the number of factors to be an arbitrary number $1 \leq k \leq k_{max}$, we estimate the factors $(\widehat{F}^k)$ using principal components method, where $k_{max} = k_{max,N,T}$ is the maximum number for possible number of factors, which is assumed to be greater or equal to the true number of factors, whose value is determined by N and T and it increases as N, T increases. Second, we select the number of factors $\hat{k}$ by minimizing a criterion modified from Bai and Ng (2002), which is a function of k and the estimated factors $(\widehat{F}^k)$. This criterion depends on the usual trade-off between good fit and parsimony. We show that this method produces a consistent estimator of the number of factors r.

The rest of this section is organized as follows. Section 2 sets up the model and presents the assumptions associated with the model. Section 3 presents the estimating procedures and the theoretical properties of the proposed estimators. Section 4 reports simulation experiments to examine the finite sample performances of our proposed method when r increases with N or T. Concluding remarks are given in Section 5. All the proofs are given in the Appendix.

3.2 Factor Models

We consider the problem of determining the number of factors (r) in a static approximate factor model, allowing $r = r_{N,T} \to \infty$, as $N \to \infty$, or $T \to \infty$, or both $N, T \to \infty$, but with a slower rate than $\min\{N, T\}$, i.e., $\max\{r/N, r/T\} \to 0$, as $N, T \to \infty$.

Let X_{it} denote the response variable for unit i at time t, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. The factor structure is of the form

$$X_{it} = \lambda_i^{0'} F_t^0 + e_{it}, \tag{3.1}$$

where F_t^0 is an $r \times 1$ vector of common factors, λ_i^0 is the $r \times 1$ vector of factor loadings, and e_{it} is the idiosyncratic component of the response variable X_{it}. There are no deterministic terms. The factors, factor loadings and idiosyncratic components are not observed. The matrix form of the factor model is

$$X = F^0 \Lambda^{0'} + e, \tag{3.2}$$

where X is a $T \times N$ matrix $(X_{ti})_{(T \times N)}$, $F^0 = (F_1, F_2, \ldots, F_T)'$ is the $T \times r$ matrix of factors, $\Lambda^0 = (\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0)'$ is the $N \times r$ matrix of factor loadings, and $e = (e_{it})_{(T \times N)}$ is the $T \times N$ matrix of idiosyncratic components.

Let $\mathrm{tr}(A)$ denote the trace of a square matrix A. The norm of matrix A is defined as $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$. Let m denote the minimum of N and T. M and C denote some generic positive constants. $\mathcal{N}$ denotes the set of natural number. We make the main assumptions as follows:

Assumption A (Factors): $\sup_{r \in \mathcal{N}} r^{-2} E\|F_t^0\|^4 < M$. Also, there exists a $r \times r$ positive definite matrix Σ_F such that $\|T^{-1} \sum_{t=1}^T F_t^0 F_t^{0'} - \Sigma_F\| \xrightarrow{p} 0$ as $T \to \infty$.

Assumption B (Factors Loadings): $\max_{1 \le i \le N} r^{-2} E\|\lambda_i^0\|^4 \le C < \infty$, and there exists a $r \times r$ positive definite matrix D such that $\|\Lambda^{0'} \Lambda^0 / N - D\| \xrightarrow{p} 0$ as $N \to \infty$.

Assumption C (Idiosyncratic Components): As $N, T \to \infty$,

1. $E(e_{it}) = 0, E|e_{it}|^8 \le M$;
2. $E(e_s' e_t) = E(N^{-1} \sum_{i=1}^N e_{is} e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \le M$ for all s, and also $T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s,t)| \le M$;
3. $E(e_{it} e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \le |\tau_{ij}|$ for some τ_{ij} and for all t; furthermore, $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \le M$;

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M$;

5. for every (t,s), $E|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M$.

6. We assume that there exist a $T\times T$ matrix L, a $N\times N$ matrix R, and a $T\times N$ matrix ε such that

$$e = L\varepsilon R$$

where L $(T\times T)$ and R $(N\times N)$ are arbitrary non-random positive definite matrices, and $\varepsilon = (\varepsilon_{ti})$ is a $T\times N$ matrix consisting of independent elements with uniformly bounded 7^{th} moment and $E(\varepsilon_{it}) = 0$.

Assumption D (Weak Dependence Between Factors and Idiosyncratic Components):

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{Tr}}\sum_{t=1}^{T}F_t^0 e_{it}\right\|^2\right) \leq M.$$

Assumptions A-B are modified from Assumptions A-B in Bai and Ng (2002). They are quite standard for factor models. Assumptions C-D are also similar to Assumptions C-D in Bai and Ng (2002), which allow for limited time-series and cross-section dependence in idiosyncratic component and also weak dependence between factors and idiosyncratic errors. Assumption C(6) puts a structure on the idiosyncratic components. This structure allows heteroskedasticity in both the time and cross-section dimensions, and also limited autocorrelation and cross-sectional correlation in the components.

3.3 Estimating the Common Factors and the Number of Factors

Following Bai and Ng (2002), we estimate the common factor in a large panel by the principal components method. For $k \in \{1, \ldots, k_{max}\}$, where k_{max} is allowed to increase at a slower speed than $\min\{N, T\}$ such that $k_{max} = o(\min\{N^{1/3}, T\})$. Let λ_i^k and F_t^k denote the loadings and factors with the allowance of k factors in the estimation. The method of principal components minimizes

$$V(k) = \min_{\Lambda^k, F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} F_t^k)^2, \tag{3.3}$$

subject to the normalization of either $\Lambda^{k\prime}\Lambda^k/N = I_k$ or $F^{k\prime}F^k/T = I_k$.

Let $\mathrm{ev}_{(i)}(A)$ denote the i^{th} largest eigenvalue of matrix A, and $\mathrm{EV}_{(i)}(A)$ is the eigenvector corresponding to the eigenvalue $\mathrm{ev}_{(\mathrm{i})}(A)$ of matrix A. If we concentrate out F^k and use the normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$, the solution to the above problem is given by $(\bar{F}^k, \bar{\Lambda}^k)$, where $\bar{\Lambda}^k = \sqrt{T}(\mathrm{EV}_{(1)}(X'X), \ldots, \mathrm{EV}_{(k)}(X'X))$. The normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$ implies $\bar{F}^k = X\bar{\Lambda}^k/N$. Define $\widehat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k)^{1/2}$, a rescaled estimator of the factors. This rescaled estimator has the asymptotic properties summarized in the following theorem.

Theorem 3.3.1 *For any* $1 \leq k \leq k_{\max} = o(\min\{N^{1/3}, T\})$ *there exists a* $(r \times k)$ *matrix* H^k *with* $\mathrm{rank} = \min\{k, r\}$ *such that*

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_t^k - H^{k\prime} F_t^0 \right\|^2 = O_p\left(\max\left\{ \frac{kr^2}{N}, \frac{k}{T} \right\} \right). \tag{3.4}$$

Similar to the results of Bai and Ng (2002), Theorem 3.3.1 suggests that the time average of the squared deviations between the estimated factors $\widehat{F}^k$ and those that lie in the true factor space, $H^{k'}F_t^0$, will vanish as $N, T \to \infty$. However, the convergence rate depends on not only the panel structure N and T, but also the factor structure r and k.

Given the results of Theorem 3.3.1, we can now analyze the problem of determining the number of factors. Let $V(k, F^k) = \min_\Lambda \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i^{k'} F_t^k)^2$ be the sum of squared residuals (divided by NT) from time-series regressions of $\underline{X}_i$ on the k factors for all $i = 1, \ldots, N$. The selecting criteria modified from those suggested by Bai and Ng (2002) have the form

$$PC(k) = V(k, \widehat{F}^k) + kg(N, T), \tag{3.5}$$

where $g(N, T)$ is the penalty factor satisfying two conditions: (i) $k_{max} \cdot g(N, T) \to 0$ as $N, T \to \infty$, (ii) $C_{N,T,k_{max}}^{-1} g(N, T) \to \infty$ as $N, T \to \infty$,where $C_{N,T,k_{max}} = O_p\left(\max\left\{\frac{k_{max}^6}{N}, \frac{k_{max}^4}{T}\right\}\right)$. As $V(k, \widehat{F}^k)$ is decreasing in k, the criteria above penalize k with a penalty factor $g(N, T)$ to select the estimator $\hat{k}$ such that under and overparameterized models will not be chosen. Theorem 3.3.2 establishes this result formally.

Theorem 3.3.2 *Let* $1 \le r \le k_{max} = o(\min\{N^{1/16}, T^{1/14}\})$. *Suppose that Assumptions A-D hold, and that (i)* $k_{max} \cdot g(N, T) \to 0$, *(ii)* $C_{N,T,k_{max}}^{-1} \cdot g(N, T) \to \infty$ *as* $N, T \to \infty$. *Denote* $\hat{k} = \text{argmin}_{1 \le k \le k_{max}} PC(k)$. *Then*

$$\lim_{N,T\to\infty} \text{Prob}[\hat{k} = r] = 1. \tag{3.6}$$

A formal proof of Theorem 3.3.2 is provided in the Appendix. Conditions (i) and (ii) together define the type of penalty factor that should vanish at an appropriate rate. They are sufficient conditions for the consistent estimation so that they may not always be required for consistent estimating the number of factors. Similar to Bai and Ng (2002), we also have the following result[1]:

Corollary 3.3.1 *Under the Assumptions of Theorem 3.3.2, if one replaces $PC(k)$ in Theorem 3.3.2 by the class of criterion defined by*

$$IC(k) = \ln\left(V(k, \widehat{F}^k)\right) + kg(N,T),$$

then the conclusion of Theorem 3.3.2 holds true.

Corollary 3.3.1 states that the class of criterion $PC(k)$ can also be used to consistently estimate the number of factors in factor models where the number of factors possibly increases with the sample sizes.

Let $\widehat{\sigma}^2$ be a consistent estimate of $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E(e_{it})^2$. Bai and Ng (2002) generalize the C_p criterion of Mallows (1973) and suggest the three PC_p criteria and as follows:

$$\begin{aligned} PC_{p1}(k) &= V(k, \widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right) \\ PC_{p2}(k) &= V(k, \widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln(\min\{N,T\}) \\ PC_{p3}(k) &= V(k, \widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{\ln(\min\{N,T\})}{\min\{N,T\}}\right) \end{aligned} \tag{3.7}$$

It is easy to check that these criteria satisfy the two conditions for the penalty factor in Theorem 3.3.2 if $k_{max} = o_p\left(\left[\ln\left(\frac{NT}{N+T}\right)\right]^{1/6}\right)$. These three criteria have different finite-sample properties while they are asymptotically equivalent. In applications,

[1]The proof of this result is omitted as it is almost the same as the proof of Corollary 1 in Bai and Ng (2002).

Bai and Ng (2002) suggest to replace $\hat{\sigma}^2$ with $V(k_{max}, \widehat{F}^{k_{max}}) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^2$, where $\hat{e}_{it} = X_{it} - \widehat{\lambda}_i^{k_{max}} \widehat{F}_t^{k_{max}}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the residuals for the linear regression of X on $\widehat{F}^{k_{max}}$. Thus, the number of factors estimated using these three criteria may be sensitive to the selection of k_{max}. Corollary 3.3.1 suggests the following three IC_p criteria can also be used to select the number of factors:

$$\begin{aligned} IC_{p1}(k) &= \ln\left(V(k, \widehat{F}^k)\right) + k \cdot \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\ IC_{p2}(k) &= \ln\left(V(k, \widehat{F}^k)\right) + k \cdot \left(\frac{N+T}{NT}\right) \ln(\min\{N,T\}), \\ IC_{p3}(k) &= \ln\left(V(k, \widehat{F}^k)\right) + k \cdot \left(\frac{\ln(\min\{N,T\})}{\min\{N,T\}}\right). \end{aligned} \tag{3.8}$$

The main advantage of these three criteria given in (3.8) is that the scaling factor $\hat{\sigma}^2$ is automatically removed by the logarithmic transformation. We do not need to estimate σ^2 before selecting the number of factors. Therefore, the number of factors estimated using IC_p criteria is insensitive to the selection of k_{max}.

As the estimated $\hat{k}$ using PC_p criteria may be sensitive to k_{max}, the selection of k_{max} is an important issue in practice. Bai and Ng (2002) suggest to select k_{max} by setting $k_{max} = 8[(\min\{N,T\}/100)^{1/4}]$ where $[A]$ denotes the integer part of a real number A. But selecting k_{max} using this rule can lead to $k_{max} < r$ since r increases with N or T in our case, which will lead to an underestimation of the number of factors because $\hat{k} \leq k_{max} < r$. On the other hand, if k_{max} is too large ($k_{max} >> r$), the selected $\hat{k}$ tend to overestimate r ($\hat{k} > r$). We propose a new procedure to resolve this problem. We propose to let k_{max} take a wide range of values. For each value of k_{max}, we select a $\hat{k}_{k_{max}}$ that minimizes the PC_p criteria. We then select the value of $\hat{k}$ that appears most times among the different $\hat{k}_{k_{max}}$ values. We use a specific example to illustrate this selection procedure. We generate a simulated data

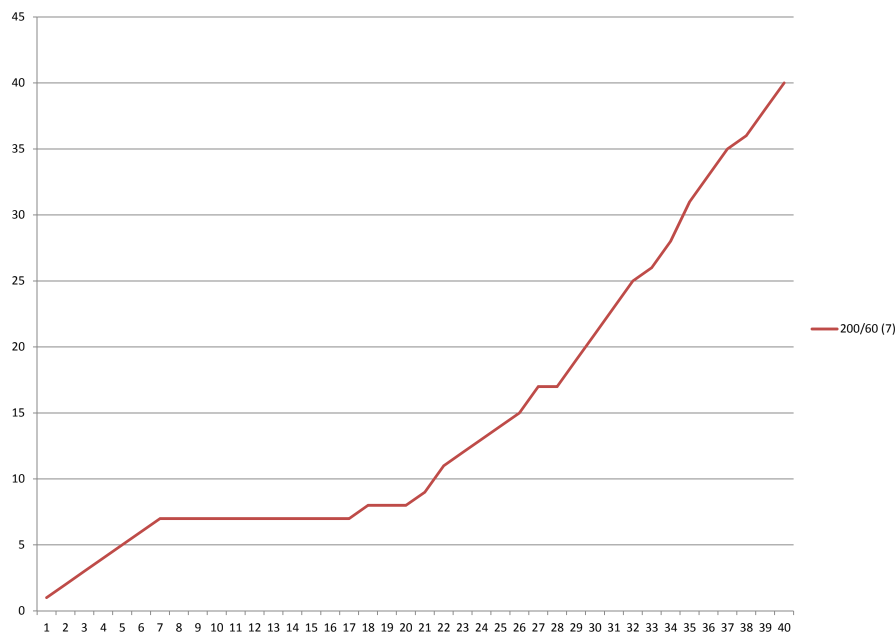

Figure 3.1: Sensitivity of PC_{p1} Criterion to k_{max}: 200/60 Case

of $N = 100$, $T = 60$ with the true number of factors $r = 7$. We let k_{max} take values from $\{1, 2, ..., 40\}$. For each different $1 \leq k_{max} \leq 40$, we select a $\hat{k}_{k_{max}}$ by minimizing PC_{p1} criterion. The result is presented in Figure 3.1. From Figure 3.1 we observe that when $k_{max} < r = 7$, we select $\hat{k} = k_{max} < 7$ as expected; when $7 \leq k_{max} \leq 16$, we select $\hat{k} = 7$; when $k_{max} > 16$, the selected $\hat{k} > 7$. Moreover, $\hat{k}$ increases with k_{max}. We also notice that $\hat{k} = 7$ is selected ten times (when $k_{max} = 7, 8, ..., 16$), while all the other values are chosen no more than three times. For example, when $17 \leq k_{max} \leq 19$, the selected $\hat{k}_{k_{max}} = 8$, i.e., $\hat{k}_{k_{max}} = 8$ is selected three times. According to our selection rule, $\hat{k} = 7$ is selected because $\hat{k} = 7$ appears most times (10 times).

Figure 3.2 plot $\hat{k}$-k_{max} curves for different N, T and r values. We see that although $\hat{k}$ increases with k_{max} for most cases, our proposed procedure can select the correct number of factors because $\hat{k}_{k_{max}}$ takes value r more often than taking any other values for all cases reported in Figure 3.2. Hence, our proposed procedure of selecting $\hat{k}$ is not sensitive to k_{max} provided that one let k_{max} take a wide range of values. Therefore, we suggest letting k_{max} to take values in $\{1, 2, ..., 40\}$ since $r \leq 40$ is likely to be true for the panel data sets economists encounter in practice.

3.4 Simulations

In this section we conduct Monte Carlo simulation to investigate how our modified criteria of Bai and Ng (2002) perform when the number of factors is allowed to increase with N or T. For simplicity of the comparison with the simulation results in Bai and Ng (2002), we first fix T and allow N and r to increase. When T is fixed as 60, we let $N = 100, 200, 500, 1000, 2000$ and $r = [1.5\log(N)]$, where $[A]$ denotes the integer part of a real number A; for $T = 100$, we let $N = 40, 60, 100, 200, 500, 1000, 2000$ and $r = [1.5\log(N)]$. The simulation results for this case are reported in the upper

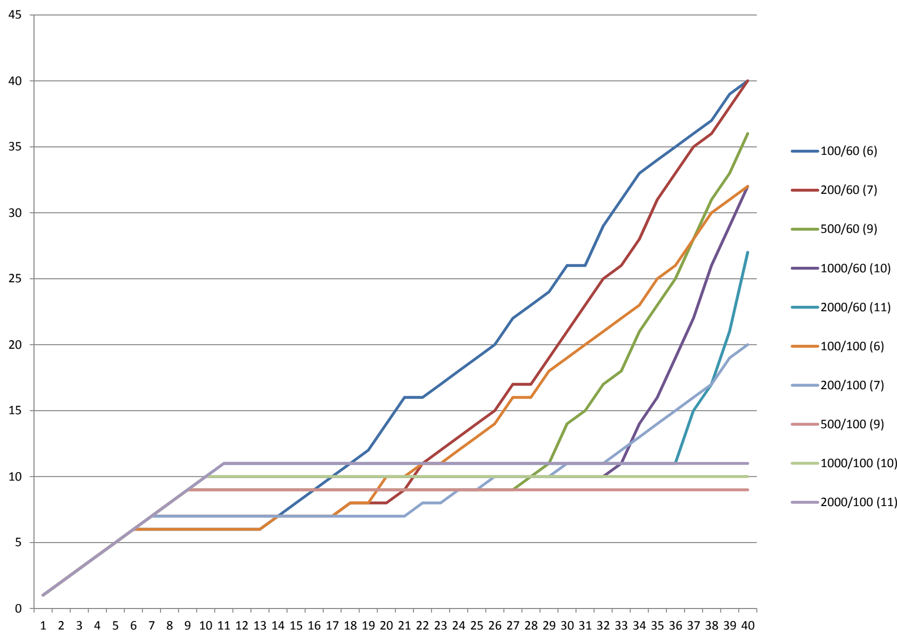

Figure 3.2: Sensitivity of PC_{p1} Criterion to k_{max}

part of each table for each data generating process (DGP). Next, we check the performance of the criteria when N is fixed and T keeps increasing. When $N = 100$, we let $T = 40$, 60 and $r = [1.5\log(T)]$; when $N = 60$, we let $T = 100, 200, 500, 1000, 2000$ and $r = [1.5\log(T)]$. The simulation results for this case are reported in the lower part of each table for each DGP. We replicate the suggested estimating procedure 1000 times and the reported results are the averages of $\hat{k}$ over 1000 replications.

The data generating processes (DGP) have the form as follows:

$$
\begin{aligned}
X_{it} =& \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}, \\
\lambda_{ij} \sim& i.i.d.N(0,1), \\
F_{tj} \sim& i.i.d.N(0,1).
\end{aligned}
$$

We consider four DGPs here. In the base case, we set the DGP as $\theta = 1$ and $e_{it} \sim i.i.d.N(0,1)$. This base DGP is denoted as DGP1. The simulation results for this case are reported in Table 3.1 where the boldfaced numbers indicate incorrect selection of the number of factors. We see that PC_{p3} selects a $\hat{k}$ larger than r in some cases. When the sample sizes are large, i.e. $\min\{N, T\} > 60$, PC_{p1}, PC_{p2}, IC_{p1} and IC_{p2} give precise estimates of the number of factors.

We denote the high-variance case as DGP2, let $\theta = 5$ and keep all the other parameters the same as those of DGP1. The estimated results of $\hat{k}$ are reported in Table 3.2. It is similar to the base case that PC_{p1}, PC_{p2} and IC_{p3} perform pretty well and give precise estimates of the number of factors for almost all cases.

For the heterogeneity case of DGP3, we set the idiosyncratic shocks to be heterogeneous. We let $\theta = 5$, and $e_{it} = u_{it} + \delta_t \epsilon_{it}$ where $u_{it} \sim i.i.d.N(0,1)$, $\epsilon_{it} \sim i.i.d.N(0,1)$, and $\delta_t = 0$ for even t, $\delta_t = 1$ for odd t. Thus the variance of the id-

Table 3.1: Estimated Number of Factors: DGP1

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	**8**	6
200	60	7	7	7	**8**	7	**8**	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	**7**	**6**	**11**	5	5	**7**
60	100	6	6	6	**12**	6	6	**9**
100	100	6	6	6	**13**	6	6	**11**
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	**6**	**6**	**10**	5	5	5
100	60	6	6	6	**12**	6	6	6
60	100	6	6	6	**12**	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP1: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; \theta = 1$, $r = [c * \ln(N)]$ *for the upper part of the table, and* $r = [c * \ln(T)]$ *for the lower part, where c=1.5, and* $[A]$ *denotes the integer part of a real number* A*.*

iosyncratic shocks is 5 when t is odd and 10 when t is even. We denote this DGP as DGP3. The estimated values of $\hat{k}$ are reported in Table 3.3. Similar to the homogeneous cases, PC_{p1} and PC_{p2} perform well when the sample sizes are large. But IC_{P1} and IC_{p2} tend to select $\hat{k}$ that is smaller than the true number of factors r, while PC_{p3} tends to overestimate r.

For the last case, denoted as DGP4, we allow the idiosyncratic to be autocorrelated. We set $\theta = 5$ and $e_{it} = \rho * e_{it-1} + v_{it}$, where $\rho = 0.5$ and $v_{it} \sim i.i.d.N(0, 1)$. The estimating results are reported in Table 3.4. The results for this case are almost the

Table 3.2: Estimated Number of Factors: High-Variance

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	6	6
200	60	7	7	7	**8**	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	**7**	5.002	**10.002**	5	5	5
60	100	6	6	6	**12**	6	6	6
100	100	6	6	6	**13**	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	10	11	11	11	11	11	11
100	40	5	5	**4**	5	**1**	**0**	5
100	60	6	**5**	**5**	6	**1**	**1**	**5**
60	100	6	6	6	**12**	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP2: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; \theta = 5,\ r = [c \ln (N)]$ *for the upper part of the table, and* $r = [c \ln (T)]$ *for the lower part, where* $[A]$ *denotes the integer part of a real number.*

same as those of the base case. When the sample sizes are large, i.e. $\min\{N, T\} > 40$, PC_{p1}, PC_{p2}, IC_{p1} and IC_{p2} perform quite well in accurately estimating the number of factors.

Summarizing the results for all the DGPs we observe that PC_{p1} and PC_{p2} have the best overall performance. Hence, we recommend using the PC_{p1} and PC_{p2} criteria in practice.

Table 3.3: Estimated Number of Factors: Heterogeneity

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6.002	6	**11.002**	**3**	**1**	**6**
200	60	7	7	7	**8**	**4**	**3**	7
500	60	9	9	9	9	**8**	**8**	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	10	10	10
40	100	5	**7**	**6**	**11**	**1**	**1**	**4**
60	100	6	6	6	**13**	**3**	**1**	**6**
100	100	6	6	6	**11**	**5**	**4**	6
200	100	7	7	7	7	**6**	**6**	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	**7**	5	**10**	**1**	**0**	**5**
100	60	6	6	6	**11**	**1**	**1**	**5**
60	100	6	6	6	**13**	**3**	**1**	**6**
60	200	7	7	7	**8**	**4**	**3**	**7**
60	500	9	9	9	9	**8**	**8**	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	10	10	10

DGP3: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; e_{it} = u_{it} + \delta_t \epsilon_{it}$, *where* $\delta_t = 0$ *for* t *even, and* $\delta_t = 1$ *for* t *odd;* $\theta = 5$, $r = [c \ln(N)]$ *for the upper part of the table, and* $r = [c \ln(T)]$ *for the lower part, where* $[A]$ *denotes the integer part of a real number.*

3.5 Concluding Remarks

In this section, we consider the problem of determining the number of factors in large factor models where the number of factors is allowed to increase, but with a slower rate, as N or T increases, i.e. $r = o(\min\{N^{1/16}, T^{1/14}\})$. We extend the analysis of Bai and Ng (2002) to the case that number of factors can increase and prove the consistency of Bai and Ng's (2002) procedure in determining the number of factors. We also propose a new procedure so that our selected number of factors is not sensitive to the selection of k_{max}. Monte Carlo simulation results suggest that

Table 3.4: Estimated Number of Factors: Autocorrelation

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	6	6
200	60	7	7	7	**8**	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	**7**	**6**	**10**	5	**3**	5
60	100	6	6	6	**12**	6	6	**7**
100	100	6	6	6	**13**	6	6	**11**
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	9	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	**7**	5	**11**	5	5	5
100	60	6	6	6	**12**	6	6	6
60	100	6	6	6	**12**	6	6	6
60	200	7	7	7	**8**	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP4: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; e_{it} = \rho e_{it-1} + v_{it}$*;* $\rho = 0.5, \theta = 5$*;* $r = [c \ln (N)]$ *for the upper part of the table, and* $r = [c \ln (T)]$ *for the lower part, where* $[A]$ *denotes the integer part of a real number.*

the criteria PC_{p1}, PC_{p2} have the overall best performance and therefore can be used to accurately estimate the number of factors when the data dimensions are relatively large, say $\min\{N, T\} \geq 60$.

One possible future research topic is to find alternative criteria that can improve the finite-sample performance of Bai and Ng's (2002) procedure and our modified procedure such that the new criteria can accurately determine the number of factors even in small or medium size samples.

4. TESTING PURCHASING POWER PARITY HYPOTHESIS: A SEMIPARAMETRIC VARYING COEFFICIENT APPROACH

Traditional linear cointegration tests of purchasing power parity (hereafter PPP) hypothesis often lead to rejection of the PPP hypothesis. More recent studies allowing for some sort of nonlinearity in econometric modelings (e.g., Michael et al. 1997) suggest mixed results and leave this problem as an unresolved issue. In this section, we analyze PPP hypothesis within a semiparametric framework using the varying coefficient model with integrated variables as considered by Cai, Li, and Park (2009), and Xiao (2009). Applying the cointegration test suggested by Xiao (2009), we conduct the cointegration test of PPP

hypothesis between U.S. and Canada, U.S. and Japan, and U.S. and U.K., respectively. In contrast to the usual findings based on linear model PPP hypothesis testing, our semiparametric model based tests support the PPP hypothesis.

4.1 Introduction

Purchasing power parity (PPP) hypothesis concerns the price level relations across countries, which is the benchmark for many theoretical and empirical analyses. There are two versions of PPP hypothesis: absolute PPP and relative PPP. Absolute PPP hypothesis suggests that the price levels between two countries should be the same once converted to the common currency. But this argument is generally not supported by empirical data. The relative law of PPP hypothesis allows for the price level to differ across countries, and states that the price behavior denominated with the common currency should be similar in the long run. Empirically, one tests the relative law of PPP hypothesis by analyzing the dynamics of exchange rate and price levels together.

Two approaches have been taken in the existing empirical investigation of PPP hypothesis. The first approach works with real exchange rate. One allows for the real exchange rate to deviate from its long run equilibrium level temporarily, but expect it to revert to its equilibrium level gradually after some shocks. So one can test whether the real exchange rate follows random walk or mean reverting to test relative law of PPP hypothesis. There are two problems with this approach. Firstly, the test suffers low power problem in finite sample applications. Meanwhile, the serial correlation coefficients are very high, which implies persistence in deviations of relative prices from PPP. Rogoff (1996) documents this phenomenon as "PPP puzzle". An alternative approach is to treat PPP as a long-run relationship, so that the deviation from PPP follows a stationary process. One can test the cointegration relations among nominal exchange rate and the price levels of the related countries to test a general version of PPP hypothesis. Culver and Papell (1999) employ this method to test the PPP hypothesis for a few countries and find cointegrating relationship. But for some economies, the data suggest rejection of cointegration relations (hence rejection to the PPP hypothesis), such as for the U.S. - Canada prices and exchange rate data. Corbae and Ouliaris (1988) find that the data of all five countries they analyze reject PPP hypothesis. There are many efforts being made trying to explain frequent rejections of the PPP hypothesis. In this section we argue that the linearity of PPP hypothesis is the main problem that leads to the rejection of PPP hypothesis test. Therefore, in this investigation we relax the assumption of linearity when testing the PPP hypothesis. We use the flexible semiparametric varying coefficient model developed by Cai, Li and Park (2009), and Xiao (2009) to re-examine the PPP hypothesis. The theoretical properties of varying coefficient model estimation with integrated variables are analyzed by Cai, Li and Park (2009), and Xiao (2009). These authors establish the consistency and derive the asymptotic distributions of their proposed estimators. Furthermore, Xiao (2009) suggests a testing procedure for testing whether the error term in the varying coefficient model follows I(0) stationary process or I(1) non-stationary process. Xiao's test can be used to test the PPP hypothesis which has the advantage of allowing for nonlinear relationship of prices between different countries. In this section we let the exchange rate to be the dependent variable and the price levels of the two countries to be the explanatory variables with varying coefficients. To model the prices relationship in the varying coefficient model framework of Cai, Li and Park (2009), and Xiao (2009), one also needs to identify a relevant stationary covariate which governs the changes of cointegration relationship between the non-stationary price variables of different countries. Since the exchange rate responds to monetary shock, we use the yield spread differentials between the two targeted countries as the additional explanatory covariate, which explains the changes in the coefficient of price level. After we finish the estimation procedure of the functional coefficient model, we compute the residuals and use them to carry out the cointegration test by following the testing procedure as suggested by Xiao (2009). We use the data of four countries (U.S., Canada, Japan, and U.K.) to form three pairs: U.S.-Canada, U.S.-Japan, and U.S.-U.K. While the traditional approach of cointegration test produces mixed results: supporting PPP hypothesis for the U.S.- U.K. pair, rejecting the PPP hypothesis for the U.S.- Canada pair, and producing an inconclusive result for U.S.- Japan data, our semiparametric varying-coefficient cointegration test offers supporting evidence of PPP hypothesis using data of all the three pairs.

Hong and Phillips (2010) have conducted similar analysis by introducing nonlinearity to PPP hypothesis. They argue that nonstationary time series have a tendency to wander without fixed mean or locality in the sample space, thus linear approximation to the possible nonliearity relationship may poorly represent the global characteristics of the processes in the long run. They suggest a modified Regression Error Specification Test (RESET) to test against both nonlinear cointegration and the absence of cointegration. Applying their modified RESET to the analysis of PPP hypothesis, Hong and Phillips (2010) find little support for a linear cointegration specification. Our work is also related to Giraitis, Kapetanios, and Yates (2014), who suggest using stochastic time-varying coefficient models to test PPP hypothesis. They use real exchange rate data to fit an auto-regression model, and focus on the mean-reversion analysis of exchange rate. Different from the above mentioned works, in this section we analyze the cointegration relationship among the nominal exchange rate and the price levels of the related countries to test the PPP hypothesis.

The rest of this section is organized as follows. Section 2 summarizes the theoretical rationale for a nonlinear PPP hypothesis, and reviews some of the existing empirical works in testing PPP hypothesis. Section 3 discusses the econometric method we use for testing the PPP hypothesis. Section 4 describes the data and compares the results of PPP hypothesis test based on a varying coefficient model with the PPP hypothesis results based traditional linear regression models. We draw conclusions and provide discussions for possible future research in section 5.

4.2 Nonlinear Exchange Rate Models

The idea of nonlinear exchange rate model dates at least back to Heckscher (1916) who considers the international transaction costs between spatially separated markets as the main cause of nonlinearity. More recently, a number of researchers (e.g., Benninga and Protopapadais 1988; Sercu et al. 1995; Michael et al. 1997; Taylor et al. 2001; O'Connell and Wei 2002; Carvalho and Nechio 2011; Chong et al. 2012) develop theoretical models of nonlinear real exchange rate adjustment and use these models to test PPP hypothesis. Based on different assumptions, one can categorize nonlinear exchange rate models into three types: transaction costs model, sticky price model, and relative productivity advantage model (Harold-Balassa-Samuelson hypothesis).

Michael et al. (1997), Baum et al. (2001), O'Connell and Wei (2002), and Imbs et al. (2003) introduce the nonlinear exchange rate adjustment model under proportional transactions costs. Proportional transaction costs create a band for exchange rate within which the marginal benefit from trade is lower than the marginal cost so that there is no arbitrage behavior. This implies the thresholds and barriers that affect the adjustment of exchange rate. And the persistent deviations from PPP are the equilibrium result of transaction costs and follows nonlinear process.

By allowing firms to price-discriminate across countries, Kollmann (2001), Atkeson and Burstein (2008), Carvalho and Nechio (2011) analyze the effect of sticky-price on the dynamics of exchange rate. They find that sticky price model generate volatile exchange rates. By allowing sectors to differ in the extent of price stickiness, Carvalho and Nechio (2011) also find heterogeneous sectoral real exchange rate dynamics. Additionally, because of price sticky, the goods markets adjust to the shocks slower than the asset markets, which implies nonlinear relationship between price levels and exchange rate.

Harrod (1933), Balassa (1964), and Samuelson (1964) observe that developed countries are relatively more productive in the traded-goods sector. As these relative productivity differences may persist over time, long-run PPP would need to be adjusted accordingly. The Harrod-Balassa-Samuelson hypothesis can also contribute to the nonlinear dynamics of exchange rate. Chong et al. (2012) check PPP hypothesis with Harrod-Balassa-Samuelson hypothesis. They compute the adjustment dynamics semiparametrically, and find that the adjustment of exchange rate toward equilibrium is relatively rapid after adjusting for productivity shocks.

Many researchers have studied empirical evidences on nonlinear exchange rate models. Michael et al. (1997) characterize the nonlinear adjustment process using a smooth transition autoregressive model, where the speed of adjustment varies with the extent of the deviation from PPP hypothesis. They show that PPP hypothesis may be rejected on the basis of a linear model even though the long-run PPP holds. Using monthly data of four countries (United States, United Kingdom, France, and Germany), they find strong evidence of mean-reverting behavior for the real exchange rate. Baum et al. (2001) confirm the results of Michael et al. (1997) using the data in the post-Bretton Woods era. They estimate exponential smooth transition autoregressive models of deviations from PPP and find clear evidence of a mean-reverting dynamic process for sizable deviations from PPP, with the equilibrium tendency varying nonlinearly with the magnitude of disequilibrium. Imbs et al. (2003) estimate various empirical models for relative prices allowing for the existence of transactions costs, and show the presence of substantial nonlinearities in exchange rate dynamics at sectoral levels.

In this analysis we use an alternative nonlinear specification, a semiparametric varying coefficient model, to test the PPP hypothesis. This flexible semiparametric specification allows for price levels from different countries and the bilateral exchange to move together with a varying nonlinear relationship. We describe the semiparametric varying coefficient model in the next section.

4.3 Varying Coefficient Cointegration Model

The PPP hypothesis is typically tested under the following setup

$$s_t = \alpha_0 + \alpha_1 p_t + \alpha_2 p_t^* + u_t \tag{4.1}$$

where s_t the logarithm of nominal exchange rate, p_t^* is the logarithm of foreign country price level, and p_t is the logarithm of home country price level. The PPP hypothesis is tested by testing the stationarity of the error term u_t. A stationary u_t process supports the PPP hypothesis, while a non-stationary I(1) process of u_t provides evidence against the PPP hypothesis. The parameters α_0, α_1 and α_2 in model (4.1) are assumed to be constants. This is a restrictive assumption as cointegration relationship may vary when economic conditions change.

Similar to the idea of linear cointegration of Granger (1981) and Engel and Granger (1987), we use cointegration to capture the long-run relationship of the economic variables. But we extend the traditional cointegration idea to a more general class which allows the cointegrating coefficients to be varying, following the approach of Xiao (2009). This framework can not only capture the long-run cointegration relationship among the exchange rate and the prices of each two countries, but also describe the nonlinear relations through varying coefficient. Thus, we can use this model to test PPP hypothesis and claim that PPP hypothesis holds if there exists cointegration relationship among the exchange rate and the prices of each two countries.

Allowing the coefficients in model (4.1) to vary with respect to the yield spread differential (z_t) leads to the following semiparametric varying coefficient model:

$$s_t = \beta_0(z_t) + \beta_1(z_t)p_t + \beta_2(z_t)p_t^* + u_t \tag{4.2}$$

where z_t is a stationary covariate that affects the relationship of p_t, p_t^* and s_t. In our empirical analysis, we choose z_t to be the yield spread differentials between the two targeted countries, because the relationship of price levels and bilateral exchange rate is likely to be affected by the change of interest rate differentials between the two countries. For the detailed definition of the yield differential variable z_t, see section 4.1. We will focus on the cointegration test of the varying coefficient model (4.2) to test PPP hypothesis.

Denote $x_t = (1, p_t, p_t^*)^T$, $\beta(z) = (\beta_0(z), \beta_1(z), \beta_2(z))^T$, we can rewrite model (4.2) as

$$s_t = x_t' \beta(z_t) + u_t \tag{4.3}$$

The local constant estimator of model (4.3) is given by

$$\widehat{\beta}_t(z) = \left[\sum_{t=1}^{T} x_t x_t^T K_h(z_t - z) \right]^{-1} \sum_{t=1}^{T} x_t s_t K_h(z_t - z) \tag{4.4}$$

where $K_h(z_t - z) = h^{-1} K\left(\frac{z_t - z}{h}\right)$, $K(\cdot)$ is the kernel function and h is the bandwidth parameter. We employ Gaussian kernel for our estimation, and choose the optimal bandwidth using least square cross validation (LS-CV) method (see Sun and Li, 2011). One can also use local linear regression to estimate model (4.3). For our analysis, these two methods produce similar estimating results.

Cai, Li, and Park (2009), and Xiao (2009) have shown the consistency and also the asymptotic distribution of this estimator.[1] We can predict the residuals using $\hat{u}_t = s_t - \hat{\beta}_0(z_t) - \hat{\beta}_1(z_t) p_t - \hat{\beta}_2(z_t) p_t^*$, and conduct the cointegration test using the method suggested by Xiao (2009). Under the null hypothesis of cointegration, u_t is a zero mean stationary process with a finite variance, i.e., $\sigma_u^2 = E(u_t^2)$ is a finite constant. If u_t follows a non-stationary unit root process, i.e. $u_t = u_{t-1} + \varepsilon_t$ with $u_0 = 0$, where ε_t is a zero mean, constant variance and serially uncorrelated (say, i.i.d.) stationary process, then $\mathrm{Var}(u_t) = t\sigma_\varepsilon^2$, which increases over time. Hence, one can test the null hypothesis of the existence of a cointegration relationship of prices

[1]Sun, Cai and Li (2013a) also consider the case that z_t is a non-stationary I(1) variable case.

and bilateral exchange rate by testing a zero coefficient of the time trend variable in the following regression model:

$$\hat{u}_t^2 = a + bt + e_t. \tag{4.5}$$

One can use the t-statistic of the coefficient of the time trend variable given by

$$\tau_T = \frac{\hat{b}}{\hat{s}(b)} \tag{4.6}$$

to test the PPP hypothesis, where $\hat{b}$ is the OLS estimator of b in model (4.5), and $\hat{s}(b) = \sqrt{\hat{\omega}^2 / \sum_{t=1}^{T} (t - \bar{t})^2}$, with $\hat{\omega}^2$ being a consistent nonparametric estimator of w^2, the long-run variance of e_t, see equation (8) of Xiao (2009) on how to obtain a consistent estimate of w^2. Xiao (2009) shows that under the null hypothesis of cointegration and some regularity conditions, this test statistic asymptotically follows standard normal distribution. We estimate the long run variance ω^2 using the methods of Andrews (1991).

4.4 Data Description and Analysis

4.4.1 Data Description

We use price-exchange rate data for four countries: U.S., Canada, Japan, and U.K. We examine whether PPP hypothesis holds between prices and exchange rate data for U.S. and Canada; U.S. and Japan, and U.S. and U.K. The three countries: Canada, Japan, and U.K. are all the largest trading partners of the United States, their economies are closely tied to the U.S. economy. The prices of Canada and U.K. exhibit almost the same dynamic pattern as that of the U.S. price (See Figure 4.1). Because of the domestic liquidity trap, Japan's price keeps decreasing after 1980s, but we can still observe simultaneous fluctuations of the prices of Japan and U.S.

Thus, it is natural to expect that the exchange rates and prices of these countries might be cointegrated with the U.S. price, i.e., PPP hypothesis may holds for the three pair countries.

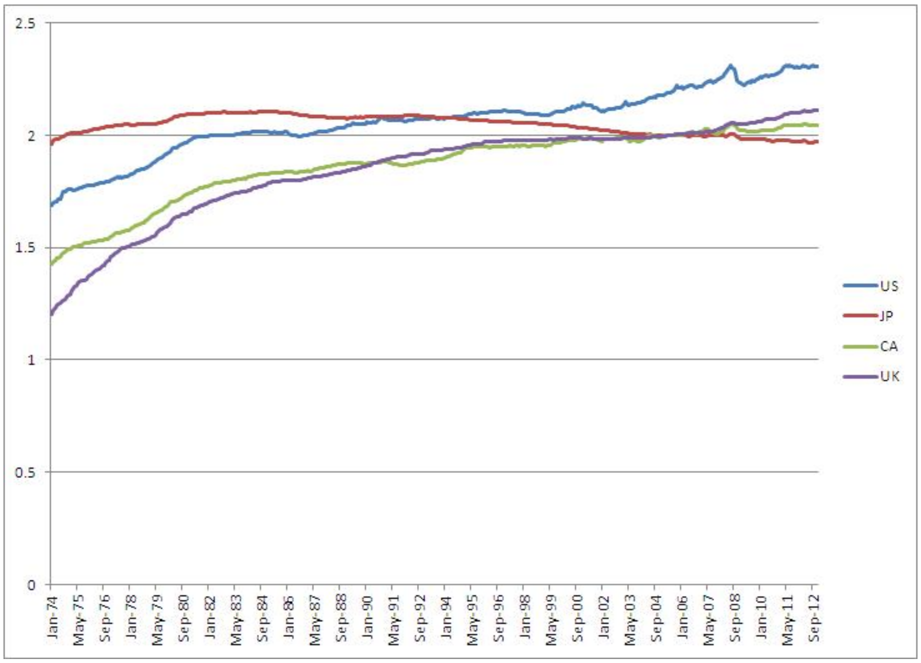

Figure 4.1: The Price Dynamics of Four Countries

For each two-country pair, we define the nominal exchange change rate as the value of one U.S. dollar ($1) in terms of the other country's currency. Thus, the nominal exchange rate for Japan means the rate at which one U.S. dollar can be exchanged for Japanese Yen. Following the previous empirical studies of PPP hypothesis test, we use producer price index (PPI) as the measure of national price level. For the stationary explanatory covariate that enters the varying coefficient function (z_t), we choose the yield spread differentials between the two targeted countries. Monthly observations on 3-month treasury bill rates ($T_{3m,t}$) and 10-year treasury bond rates ($T_{10y,t}$) for all the countries are used to construct the yield spread. Thus, the covariate z_t is defined as $z_t \equiv (T_{10y,t,i} - T_{3m,t,i}) - (T_{10y,t,US} - T_{3m,t,US})$, with $i = CA, JP,$ and UK for Canada, Japan, and U.K., respectively. Monthly data are taken from Datastream offered by Thomson Reuters. The data span the period from 1974m1 to 2013m1.

4.4.2 Traditional Cointegration Analysis

We use augmented Dickey-Fuller (ADF) test, the GLS modified ADF (DF-GLS) test of Elliott, Rothenberg, and Stock (1996), and Phillips and Perron's (1988, or PP) z_α test to test for unit roots for each price and exchange rate variable. The results are summarized in Table 4.1. The ADF test for all the time series except p_{CA} or p_{UK} fail to reject the unit root hypothesis. Although the ADF test for p_{CA} and p_{UK} suggest these two time series are stationary, the DF-GLS and PP tests fail to reject the unit root hypothesis. Thus, we conclude that p_{CA} and p_{UK} might be nonstationary as well. The overall testing results suggest that all the price levels and the exchange rates are likely to be I(1) processes in keeping with the conventional wisdom.

Table 4.1: Unit Root Tests of the Time Series

Variables	ADF	DF-GLS	PP
p_{US}	-2.172	-0.776	-9.582
s_{CA}	-1.519	-0.726	-3.518
p_{CA}	-4.703^{***}	0.128	-5.258
s_{JP}	-1.214	-2.409	-10.511
p_{JP}	-1.079	-0.059	-7.815
s_{UK}	-1.955	-3.373^{**}	-15.676^{**}
p_{UK}	-3.556^{***}	-0.705	-6.529

Notes: ***,**,and * denote rejecting the null hypothesis of unit root at 1%, 5%, and 10% significance levels, respectively.

We test the potential cointegrating relationship among the three integrated variables using two cointegration tests: Engle-Granger (EG) residual-based cointegrating test and Johansen's cointegrating rank test. The statistics are summarized in Table 4.2. The EG cointegrating test to model (4.1) produce a t-statistic equal to -2.493 with a p-value of 0.1172 for the U.S.-Canada pair, a t-statistic equal to -2.430 with a p-value of 0.1335 for the U.S.-Japan pair, indicating failure to reject a spurious regression at conventional significance levels. But the t-statistic of U.S. - U.K. pair is -3.3372 with a p-value of 0.0120 using EG cointegrating test, which rejects the null hypothesis of spurious regression at 5% level.

Johansen's cointegrating rank test statistic of U.S. - Canada pair for the null hypothesis of no cointegrating relation against the alternative hypothesis of at least one cointegrating relation yields a value of 34.1433, which is smaller than the 5% critical value of 34.55. This confirms the result of EG residual-based cointegrating test, suggesting no cointegration for U.S. - Canada pair. For U.S. - Japan pair, Johansen's cointegrating rank test suggests rank one cointegrating relation, different from the result of Engle-Granger (EG) residual-based cointegrating test. This result shares similar pattern of Kugler and Lenz's (1993) work, which concludes that Johansen's test is more favorable to PPP hypothesis in the long run than Engel-Granger approach. For U.S. - U.K. pair, Johansen's cointegrating rank test produces the same conclusion as the result of EG cointegrating test.

Thus, traditional method of cointegration test for PPP hypothesis suggests mixed results. For U.S. - Canada pair, the general version of PPP hypothesis is rejected while it is supported by the data of U.S. - U.K. pair with the traditional method of cointegration test, consistent with the existing literature.

Meanwhile, we test the null hypothesis that the model is linear for the three pairs using the method of Sun, Cai and Li (2013b). We find that each null hypothesis is

Table 4.2: Cointegration Test of Each Pair

Pair	Engle-Granger		Johansen	
U.S. - Canada	-2.493	(0.1172)	34.1433	(H_0: rank=0)
U.S. - Japan	-2.430	(0.1335)	14.6439	(H_0: rank=1)
U.S. - U.K.	-3.372^{**}	(0.0120)	0.9592	(H_0: rank=2)

rejected at the 0.1% level, which also suggests that we should consider nonlinear model to test PPP hypothesis.

4.4.3 Semiparametric Results

Next, we use the semiparametric varying coefficient cointegration model to re-examine the PPP hypothesis. When applying the ADF test and the PP test to the yield difference z_t for all three pairs, we reject the null hypothesis of a unit root at the 5% significance level (a linear time trend and one lag are included in the ADF test). Therefore, the assumption that z_t is stationary is not rejected at the 5% significance level.

Applying the least squares cross-validation method of Sun and Li (2011) to model (4.2), we obtain the bandwidths for each pair: $h^*_{CA} = 0.1808203$, $h^*_{JP} = 0.2193756$, and $h^*_{UK} = 0.8089695$. It is known that the least squares cross validation (LS-CV) method can sometimes select a bandwidth value too small (under-small) and lead to a wiggly fitted curve. This indeed happens for the U.S. - Japan data case. We therefore multiply the LS-CV bandwidth h^*_{JP} by a factor of 1.5 to get the bandwidth for our regression $h_{JP} = 0.3290634$. The estimated curves are plotted in Figures 4.2, 4.3 and 4.4. As the strong version of PPP hypothesis suggest that the coefficient for p_t, β_2 is equal to 1, and the coefficient for p^*_t, β_1 is equal to -1, we plot the semiparamtric estimates of $-\beta_1(\cdot)$ and $\beta_2(\cdot)$. Figure 4.2, 4.3 and 4.4 show the result of U.S. - Canada pair, U.S. - Japan pair, and U.S. - U.K. pair accordingly. We can find that the estimated values of $\beta_2(z)$ are all positive, while the estimated values $\beta_1(z)$ are all negative for all the three pairs (so that $-\beta_1(z)$ is positive), which is consistent with the theoretic prediction of PPP hypothesis. Additionally, the estimated coefficient curves of $-\beta_1(z)$ and $\beta_2(z)$ exhibit similar shapes and values for all the three pairs, which is consistent with the theoretical prediction of PPP hypothesis as well.

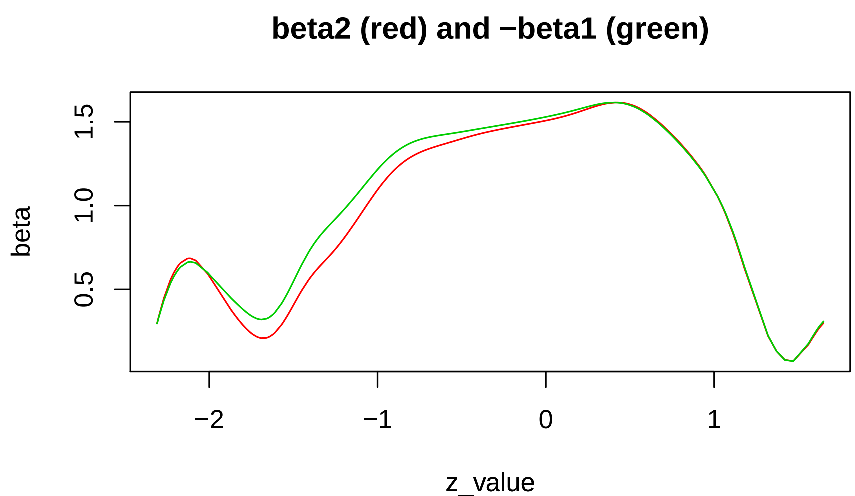

Figure 4.2: The Estimated Coefficients for Canada

The t-statistics of functional-coefficient cointegration test suggested by Xiao (2009) are summarized in Table 4.3. The value of the t-statistic for U.S. - Canada pair is -1.108889, for U.S. - Japan pair is -0.7585207, and for U.S. - U.K. pair is -1.182685. All of them are not significant at 10% significance level, thus fail to reject the null hypothesis that the square of the estimated residuals has no time trend. This suggests the cointegration relationship holds for all the three pairs. Therefore, semiparametric functional-coefficient cointegration test offers strong supporting evidence of PPP

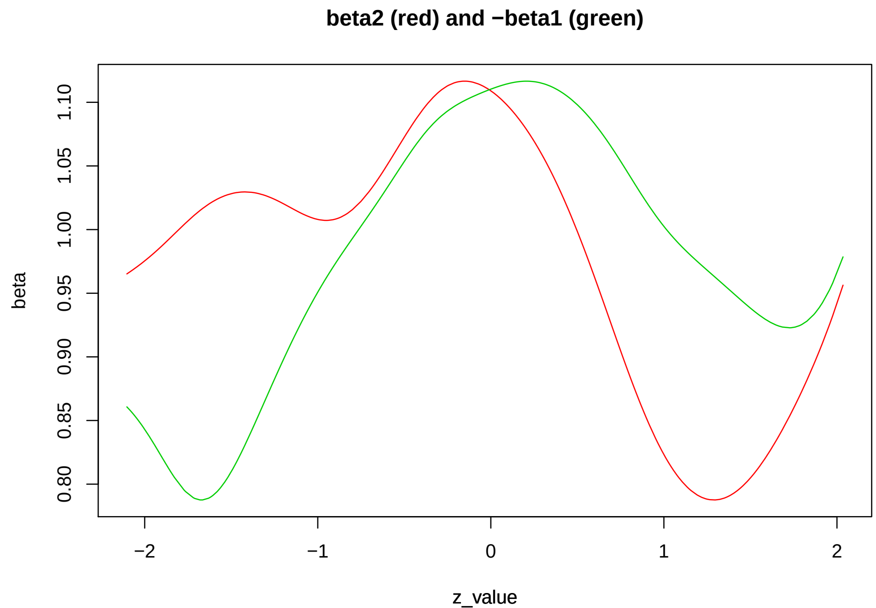

Figure 4.3: The Estimated Coefficients for Japan

hypothesis, even when the PPP hypothesis is rejected by the traditional approach of linear model based cointegration test.

Table 4.3: Functional-Coefficient Cointegration Test of Each Pair

	U.S. - Canada	U.S. - Japan	U.S. - U.K.
bandwidth	0.1808203	0.3290634	0.8089695
t-statistic	-1.108889	-0.7451862	-1.191916

4.5 Concluding Remarks

This study revisits the highly controversial issue of PPP hypothesis using semi-parametric approach with monthly data of U.S. - Canada, U.S. - Japan, and U.S. - U.K. pairs. Our findings demonstrate that, while the traditional cointegration test

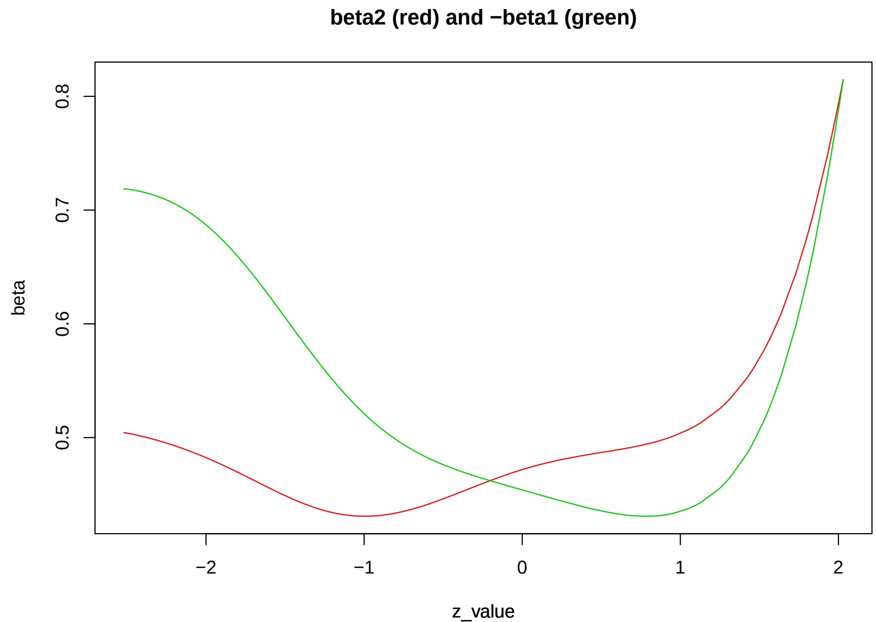

Figure 4.4: The Estimated Coefficients for U.K.

tends to reject PPP hypothesis, our semiparametric functional-coefficient approach provides supporting evidence of the general version of PPP hypothesis.

In this work we do not consider the possible aggregation bias and measurement error problems. Imbs *et al.* (2005) argue that the aggregation bias may cause rejection of the PPP hypothesis. They suggest that when one has a bunch of AR processes, the aggregation of them is going to behave closely to the most persistent one. In contrast, Chen and Engel (2005) argue that the aggregation bias is quite small and error-in-variables in the data can make price series appear less persistent than they actually are. Future researches can be conducted by modeling the price-exchange relationship, allowing for both the aggregation bias and error-in-variable, within the semiparametric framework to test the PPP hypothesis.

5. SUMMARY

This dissertation considers three questions related to the theory and application of semiparametric methods: the optimal bandwidth selection of CDF estimation of mixed data-type multivariate covariates, the determination of number of factors for factor models where the number of factors is allowed to increase with sample sizes, and the application of varying coefficient models to the test of "purchasing power parity" hypothesis.

For the problem of bandwidth selection, I provide a cross-validated approach to select the optimal bandwidths for kernel estimation of CDF functions that admits both discrete and continuous variables. I prove that the estimator of CDF functions converges to normal distribution asymptotically at the rate of $\sqrt{n}$. An application to estimation of a bivariate copula demonstrates the potential utility of the proposed method for practitioners. The approach has the added benefit of being computationally appealing due to the manner in which the cross-validation function can be expressed, while an implementation in R (R Core Team, 2013) is available in the np package.

To determine the number of factors for factor models, I propose to estimate the number of factors which is allowed to increase with sample sizes using the information criteria. In my analysis, the number of factors r can increase at a relatively slower rate as N or T increases, i.e. $r = o(\min\{N^{1/16}, T^{1/14}\})$. I extend the analysis of Bai and Ng (2002) and show the consistency of Bai and Ng's (2002) method in determining the number of factors. I also propose a new procedure so that the selected number of factors is not sensitive to the selection of k_{max}. Monte Carlo simulation results suggest that the criteria PC_{p1} and PC_{p2} have the overall best performance and therefore can be used to accurately estimate the number of factors when the data dimensions are relatively large, say $\min\{N, T\} \geq 60$.

To apply the varying coefficient models to the test the "purchasing power parity" hypothesis, I analyze the monthly data of U.S. - Canada, U.S. - Japan, and U.S. - U.K. pairs. I find that the semiparametric functional-coefficient approach provides supporting evidence of the general version of PPP hypothesis while the traditional cointegration test tends to reject PPP hypothesis. This also shows how semiparametric/nonparametric methods can alleviate the problem of model misspecification.

As a whole, these three essays provide insights into the theoretical properties and empirical applications of semiparametric econometrics. The discussions of bandwidth selection for CDF estimation can be applied to copula models, quantile estimation and survival analysis. The results of the determination of the number of factors for factor models can be used for financial analysis, macroeconomic forecasting and policy evaluation. Also, the "purchasing power parity" test using varying coefficient model shows one of the applications of semiparametric/nonparametric methods for empirical analyses.

REFERENCES

Andrews, D.W. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.

Anh, S. and Horenstein, A. (2013) Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203-1127.

Atkeson, A. and Burstein, A. (2008) Pricing-to-market, trade costs, and international relative prices. *American Economic Review* 98, 1998-2031.

Bai, J. and Ng, S. (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191-221.

Balassa, B. (1964) The purchasing power parity doctrine: a reappraisal. *Journal of Political Economy* 72, 584-596.

Bashtannyk, D. M. and Hyndman, R. J. (2001) Bandwidth selection for kernel conditional density estimation. *Computational Statistics and Data Analysis* 36, 279-298.

Baum, C.F., Barkoulas, J.T. and Caglayan, M. (2001) Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. *Journal of International Money and Finance* 20, 379-399.

Benninga, S. and Protopapadakis, A. (1988) The equilibrium pricing of exchange rates and assets when trade takes time. *Journal of International Money and Finance* 7, 129-149.

Bowman, A. W., Hall, P. and Prvan, T. (1998) Bandwidth selection for the smoothing of distribution functions. *Biometrika* 85, 799-808.

Cai, Z., Li, Q. and Park, Y. (2009) Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101-113.

Carvalho, C. and Nechio, F. (2011) Aggregation and the PPP puzzle in a sticky-price model. *American Economic Review* 101, 2391-2424.

Chen, S. and Engel, C. (2005) Does 'aggregation bias' explain the PPP puzzle? *Pacific Economic Review* 10, 49-72.

Chen, S. X. and Huang, T. (2007) Nonparametric estimation of copula functions. *Canadian Journal of Statistics* 35, 265-282.

Chong, Y., Jordá, Ó. and Taylor, A.M. (2012) The Harrod-Balassa-Samuelson hypothesis: real exchange rates and their long-run equilibrium. *International Economic Review* 53, 609-634.

Connor, G. and Korajzcyk, R. (1993) A test for the number of factors in an approximate factor model. *Journal of Finance* 48, 1263-1291.

Corbae, D. and Ouliaris, S. (1988) Cointegration and tests of purchasing power parity. *Review of Economics and Statistics* 70, 508-511.

Cragg, J. and Donald, S. (1997) Inferring the rank of a matrix. *Journal of Econometrics* 76, 223-250.

Culver, S.E. and Papell, D.H. (1999) Long-run purchasing power parity with short-run data: evidence with a null hypothesis of stationarity. *Journal of International*

Money and Finance 18, 751-768.

Donald, S. (1997) Inference concerning the number of factors in a multivariate nonparameteric relationship. *Econometrica* 65, 103-132.

Elliott, G., Rothenberg, T.J. and Stock J.H. (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813-836.

Engel, R.F. and Granger, C.W.J. (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.

Fan, J., Liao, Y. and Mincheva, M. (2011) High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics* 39, 3320-3356.

Fan, J., Liao, Y. and Mincheva, M. (2013) Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B* 75, 603-680.

Forni, M., and Gambetti, L. (2010) Macroeconomic shocks and the business cycle: evidence from a structural factor model. *Working Papers 440*, Barcelona Graduate School of Economics.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.(2000) Reference cycles: the NBER methodology revisited. *CEPR Discussion Paper 2400*.

Forni, M., and Reichlin, L. (1998) Let's get real: a factor-analytic approach to disaggregated business cycle dynamics. *Review of Economic Studies* 65, 453-473.

Giannone, D., Reichlin, L. and Sala, L. (2005) Monetary policy in real time. *NBER Macroeconomics Annual 2004*, e.d. by Mark Gertler and Kenneth Rogoff, 161-200.

MIT Press.

Giraitis, L., Kapetanios, G. and Yates, T. (2014) Inference on stochastic time-varying coefficient models. *Journal of Econometrics* 179, 46-65.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121-123.

Gregory, A. and Head, A. (1999) Common and country-specific fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics*, 44, 423-452.

Hall, P., Racine, J. and Li, Q. (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99, 1015-1026.

Harrod, R.F. (1933) *International Economics*, Nisbet and Cambridge University Press, London.

Hayfield, T. and Racine, J.(2008) Nonparametric econometrics: the np package. *Journal of Statistical Software* 27. URL: http://www.jstatsoft.org/v27/i05/

Heckscher, E.F. (1916) Växelkursens grundval vid pappersmyntfot. *Ekono-misk Tidskrift* 18, 209-312

Hong, S.H. and Phillips, P.C.B. (2010) Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business and Economic Statistics* 28, 96-114.

Hsiao, C., Ching, H. S. and Wan, S. K. (2012) A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with Mainland China. *Journal of Applied Econometrics* 27, 705-740.

Imbs, J., Mumtaz, H., Raven, M. O. and Rey, H. (2003) Nonlinearities and real exchange rate dynamics. *Journal of the European Economic Association* 1, 639-649.

Imbs, J., Mumtaz, H., Raven, M. O. and Rey, H. (2005) PPP strikes back: aggregation and the real exchange rate. *Quarterly Journal of Economics* 120, 1-43.

Jurado, K., Ludvigson, S. and Ng, S. (2013) Measuring uncertainty. *NBER Working Paper No. 19456*.

Kollmann, R. (2001) The exchange rate in a dynamic-optimizing business cycle model with nominal rigidities: a quantitative investigation. *Journal of International Economics* 55, 243-262.

Kugler, P. and Lenz, C. (1993) Multivariate cointegration analysis and the long-run validity of PPP. *Review of Economics and Statistics* 75, 180-184.

Lewbel, A. (1991) The rank of demand systems: theory and nonparametric estimation. *Econometrica* 59, 711-730.

Li, Q., Lin, J. and Racine, J. (2013) Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business and Economic Statistics* 31(1), 57-65.

Li, Q. and Racine, J. (2003) Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86(2), 266-292.

Li, Q. and Racine, J. (2008) Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics* 26(4), 423-434.

Liu, R. and Yang, L. (2008) Kernel estimation of multivariate cumulative distribution function. *Journal of Nonparametric Statistics* 20(8), 661-677.

Ludvigson, S. and Ng, S. (2007) The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics* 83, 171-222.

Ludvigson, S. and Ng, S. (2009) Macro factors in bond risk premia. *Review of Financial Studies* 22, 5027-5067.

Ludvigson, S. and Ng, S. (2010) A factor analysis of bond risk premia. *Handbook of Empirical Economics and Finance*, e.d. by Aman Uhla and David E. A. Giles, 313-372. Chapman and Hall, Boca Raton, FL.

Mallows, C. L. (1973) Some comments on C_p. *Technometrics* 15, 661-675.

Michael, P., Nobay, R. and Peel D. (1997) Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy* 105, 862-879.

Nelsen, R. B. (2006) *An Introduction to Copulas*, second edn, Springer, New York.

O'Connell, P.G.J., and Wei, S.-J. (2002) The bigger they are, the harder they fall: the retail price differences across U.S. cities. *Journal of International Economics* 56, 21-54.

Onatski, A. (2009) Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447-1479.

Onatski, A. (2010) Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92, 1004-1016.

Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in a time series regression. *Biometrika* 75, 335-346.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna. URL: http://www.R-project.org/

Rogoff, K. (1996) The purchasing power parity puzzle. *Journal of Economic Literature* 34, 647-668.

Ross, S. (1976) The arbitrage theory of capital asset pricing. *Journal of Finance* 13, 341-360.

Samuelson, P.A. (1964) Theoretical notes on trade problems. *Review of Economics and Statistics* 46, 145-154.

Sercu, P., Uppal, R. and Van Hulle, C. (1995) The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity. *Journal of Finance* 50, 1309-1319.

Stock, J. H. and Watson, M. (1998) Diffusion indexes. *NBER Working Paper 6702.*

Stock, J. H. and Watson, M. (1999) Forecasting inflation. *Journal of Monetary Economics* 44, 293-335.

Sun, Y., Cai, Z. and Li, Q. (2013a) Semiparametric functional coefficient models with integrated covariates. *Econometric Theory* 79, 659-672.

Sun, Y., Cai, Z. and Li, Q. (2013b) Consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. *Working paper*, University of Guelph.

Sun, Y. and Li, Q. (2011) Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business and Economic Statistics* 29, 541-551.

Taylor, M.P., Peel, D.A. and Sarno, L. (2001) Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles. *International Economic Review* 42, 1015-1042.

Xiao, Z. (2009) Functional-coefficient cointegrating regression. *Journal of Econometrics* 152, 81-92.

APPENDIX A

PROOFS OF THEOREMS IN SECTION 2

Since the expected value of the estimator $\hat{F}(y)$ is

$$E\hat{F}(y) = E\left\{\frac{1}{n}\sum_{j=1}^{n} G_\gamma(y, y_j)\right\} = EG_\gamma(y, y_j)$$

we can have the bias of the estimator $\hat{F}(y)$ as

$$\text{bias}\left(\hat{F}(y)\right) = EG_\gamma(y, y_j) - F(y) \tag{A.1}$$

Also, we have

$$\text{Var}\left(\hat{F}(y)\right) = \text{Var}\left\{\tfrac{1}{n}\textstyle\sum_{j=1}^{n} G_\gamma(y, y_j)\right\} = \tfrac{1}{n}\text{Var}\big(G_\gamma(y, y_j)\big)$$

So we have the mean square error(MSE) of $\hat{F}(y)$ as

$$\text{MSE}\left(\hat{F}(y)\right) = \frac{1}{n}\text{Var}\big(G_\gamma(y, y_j)\big) + \big(EG_\gamma(y, y_j) - F(y)\big)^2 \tag{A.2}$$

We choose bandwidths by minimizing the cross-validation function

$$CV(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\int \{I_i - \hat{F}_{-i}\}^2 dy$$

where I_i denotes $I(y_i \leq y)$, and $\hat{F}_{-i} = \frac{1}{n-1}\sum_{j\neq i}^{n} G_\gamma(y, y_j)$ for each $i = 1, \cdots, n$, the leave-one-out estimator of $F(y)$.

Proof of Lemma 2.3.1

We have

$$
\begin{aligned}
& G_\gamma(y, y_j) \\
= & \sum_{z^d \le y^d} \lambda_1^{|y_{1j}^d - z_1^d|} \cdots \lambda_r^{|y_{rj}^d - z_r^d|} \int_{-\infty}^{y^c} \frac{1}{h_1} w\left(\frac{z_1^c - y_{1j}^c}{h_1}\right) \cdots \frac{1}{h_q} w\left(\frac{z_q^c - y_{jq}^c}{h_q}\right) dz_1^c \cdots dz_q^c \\
\equiv & \sum_{z^d \le y^d} \lambda^{|y_j^d - z^d|} G\left(\frac{y^c - y_j^c}{h}\right) \qquad \text{(A.3)}
\end{aligned}
$$

where $G\left(\frac{y^c - y_j^c}{h}\right)$ denotes $\int_{-\infty}^{y_1^c} \cdots \int_{-\infty}^{y_q^c} \frac{1}{h_1} w\left(\frac{z_1^c - y_{1j}^c}{h_1}\right) \cdots \frac{1}{h_q} w\left(\frac{z_q^c - y_{jq}^c}{h_q}\right) dz_1^c \cdots dz_q^c$, and $\lambda^{|y_j^d - z^d|}$ denotes $\lambda_1^{|y_{1j}^d - z_1^d|} \cdots \lambda_r^{|y_{rj}^d - z_r^d|}$.

The expectation of $G_\gamma(y, y_j)$ is

$$
\begin{aligned}
E_{y_j} G_\gamma(y, y_j) &= E_{y_j} \left\{ \sum_{z^d \le y^d} \lambda^{|y_j^d - z^d|} G\left(\frac{z^c - y_j^c}{h}\right) \right\} \\
&= E_{y_j^d} \left\{ \sum_{z^d \le y^d} \lambda^{|y_j^d - z^d|} E_{y_j^c | y_j^d} \left\{ G\left(\frac{z^c - y_j^c}{h}\right) \right\} \right\} \qquad \text{(A.4)}
\end{aligned}
$$

where

$$
\begin{aligned}
& E_{y_j^c | y_j^d} \left\{ G\left(\frac{z^c - y_j^c}{h}\right) \right\} \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G\left(\frac{z^c - y_j^c}{h}\right) f(y_{1j}^c, \cdots, y_{qj}^c | y_j^d) dy_{1j}^c \cdots dy_{qj}^c \\
= & F_{Y^c | Y^d}(y^c | y_j^d) + \frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 F_{ss}(y^c | y_j^d) + O\left(\sum_{s=1}^{q} h_s^4\right)
\end{aligned}
$$

Define

$$
d(y, z) = \begin{cases} 1, & \textit{if } |y - z| = 1; \\ 0, & \textit{otherwise.} \end{cases}
$$

Then

$$\lambda_r^{|y_{rj}^d - z_r^d|} = \mathbf{1}(y_{rj}^d = z_r^d) + \lambda_s d(y_{rj}^d, z_r^d) + O(\lambda_s^2)$$

Let $\mathbf{1}(y_j^d = z^d)$ denote $\prod_{s=1}^r \mathbf{1}(y_{sj}^d = z_s^d)$ and $\mathbf{1}(y_{-sj}^d = z_{-s}^d)$ denote $\prod_{l\neq s}^r \mathbf{1}(y_{lj}^d = z_l^d)$, then we have

$$\begin{aligned}
& E_{y_j} G_\gamma(y, y_j) \\
= \; & E_{y_j^d}\Bigg\{ \sum_{z^d \leq y^d} \Bigg[\big(\mathbf{1}(y_{rj}^d = z_r^d) + \lambda_s d(y_{rj}^d, z_r^d)\big) \cdots \big(\mathbf{1}(y_{rj}^d = z_r^d) + \lambda_s d(y_{rj}^d, z_r^d)\big) \\
& + O\left(\sum_{s=1}^r \lambda_s^2\right)\Bigg] \times \left[F_{Y^c|Y^d}(y^c|y_j^d) + \frac{\kappa_2}{2}\sum_{s=1}^q h_s^2 F_{ss}(y^c|y_j^d) + O\left(\sum_{s=1}^q h_s^4\right)\right]\Bigg\} \\
= \; & F(y^c, y^d) + \frac{\kappa_2}{2}\sum_{s=1}^q h_s^2 F_{ss}(y^c, y^d) \\
& + \sum_{s=1}^r \lambda_s E_{y_j^d}\left\{\sum_{z^d \leq y^d} \mathbf{1}(y_{-sj}^d = z_{-s}^d) d(y_{sj}^d, z_s^d) F(y^c|y_j^d)\right\} \\
& + O\left(\sum_{s=1}^r \lambda_s^2 + \sum_{s=1}^q h_s^4\right) \\
\equiv \; & F(y^c, y^d) + \frac{\kappa_2}{2}\sum_{s=1}^q h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^r \lambda_s B_{1s}(y^c, y^d) + O\left(\sum_{s=1}^r \lambda_s^2 + \sum_{s=1}^q h_s^4\right)
\end{aligned}$$

where $B_{1s}(y^c, y^d) \equiv E_{y_j^d}\left\{\sum_{z^d \leq y^d} \mathbf{1}(y_{-sj}^d = z_{-s}^d) d(y_{sj}^d, z_s^d) F(y^c|y_j^d)\right\}$.

Similarly, we can calculate the expectation of $G^2(y, y_j)$ as

$$\begin{aligned}
E_{y_j} G_\gamma^2(y, y_j) &= E_{y_j}\left\{\sum_{z^d \leq y^d} \lambda^{2|y_j^d - z^d|} G^2\left(\frac{z^c - y_j^c}{h}\right)\right\} \\
&= E_{y_j^d}\left\{\sum_{z^d \leq y^d} \lambda^{2|y_j^d - z^d|} E_{y_j^c|y_j^d}\left\{G^2\left(\frac{z^c - y_j^c}{h}\right)\right\}\right\}
\end{aligned}$$

where

$$
\begin{aligned}
& E_{y_j^c|y_j^d}\left\{G^2\left(\frac{z^c-y_j^c}{h}\right)\right\} \\
= \;& \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} G^2\left(\frac{z^c-y_j^c}{h}\right) f(y_{1j}^c,\cdots,y_{qj}^c|y_j^d)dy_{1j}^c\cdots dy_{qj}^c \\
= \;& F_{Y^c|Y^d}(y^c|y_j^d) - \alpha_0\sum_{s=1}^q h_s F_s(y^c|y_j^d) + O\left(\sum_{s=1}^q h_s^2\right)
\end{aligned}
$$

where $\alpha_0 = 2\int vG(v)w(v)dv$.

Thus, we have

$$
\begin{aligned}
& E_{y_j} G_\gamma^2(y, y_j) \\
= \;& E_{y_j^d}\Bigg\{\Bigg[\sum_{z^d\le y^d}\left(\mathbf{1}(y_j^d = z^d)\right) + 2\lambda_1\sum_{z^d\le y^d}\left(\mathbf{1}(y_j^d = z^d)\right)d(y_{1j}^d, z_1^d) + \cdots \\
& +2\lambda_r\sum_{z^d\le y^d}\left(\mathbf{1}(y_j^d = z^d)\right)d(y_{rj}^d, z_r^d) + O\left(\sum_{s=1}^r \lambda_s^2\right)\Bigg] \\
& \left[F_{Y^c|Y^d}(y^c|y_j^d) - \alpha_0\sum_{s=1}^q h_s F_s(y^c|y_j^d) + O\left(\sum_{s=1}^q h_s^2\right)\right]\Bigg\} \\
= \;& F(y^c, y^d) - \alpha_0\sum_{s=1}^q h_s F_s(y^c, y^d) \\
& +2\sum_{s=1}^r \lambda_s E_{y_j^d}\left\{\sum_{z_d\le y^d}\mathbf{1}(y_j^d = z^d)d(y_{sj}^d, z_s^d)F(y^c|y_j^d)\right\} \\
& +O\left(\sum_{s=1}^r \lambda_s^2 + \sum_{s=1}^q h_s^2\right) \\
\equiv \;& F(y^c, y^d) - \alpha_0\sum_{s=1}^q h_s F_s(y^c, y^d) + 2\sum_{s=1}^r \lambda_s B_{2s}(y^c, y^d) \\
& +O\left(\sum_{s=1}^r \lambda_s^2 + \sum_{s=1}^q h_s^2\right)
\end{aligned}
$$

where $B_{2s}(y^c, y^d) \equiv E_{y_j^d}\left\{\sum_{z_d\le y^d}\mathbf{1}(y_j^d = z^d)d(y_{sj}^d, z_s^d)F(y^c|y_j^d)\right\}$.

And the variance of $G_\gamma(y, y_j)$ is

$$\begin{aligned}
\mathrm{Var}(G_\gamma(y, y_j)) &= E_{y_j} G_\gamma^2(y, y_j) - \left[E_{y_j} G_\gamma(y, y_j)\right]^2 \\
&= F(y^c, y^d) - \alpha_0 \sum_{s=1}^{q} h_s F_s(y^c, y^d) + 2 \sum_{s=1}^{r} \lambda_s B_{2s}(y^c, y^d) \\
&\quad - \Bigg[F(y^c, y^d) + \frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s B_{1s}(y^c, y^d) \\
&\quad + O\left(\sum_{s=1}^{r} \lambda_s^2 + \sum_{s=1}^{q} h_s^4 \right) \Bigg]^2 + O\left(\sum_{s=1}^{r} \lambda_s^2 + \sum_{s=1}^{q} h_s^2 \right) \\
&= F(y)(1 - F(y)) - \alpha_0 \sum_{s=1}^{q} h_s F_s(y^c, y^d) \\
&\quad + 2 \sum_{s=1}^{r} \lambda_s \left[B_{2s}(y^c, y^d) - F(y) B_{1s}(y^c, y^d) \right] \\
&\quad + O\left(\sum_{s=1}^{r} \lambda_s^2 + \sum_{s=1}^{q} h_s^2 \right) \qquad \text{(A.5)}
\end{aligned}$$

Therefore, the MSE of $\hat{F}(y)$ is

$$\begin{aligned}
\mathrm{MSE}\left(\hat{F}(y)\right) &= \frac{1}{n} \mathrm{Var}(G_\gamma(y, y_j)) + \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \\
&= \frac{1}{n} F(y)(1 - F(y)) - \alpha_0 \sum_{s=1}^{q} \frac{h_s}{n} F_s(y^c, y^d) \\
&\quad + 2 \sum_{s=1}^{r} \frac{\lambda_s}{n} \left[B_{2s}(y^c, y^d) - F(y) B_{1s}(y^c, y^d) \right] \\
&\quad + \left(\frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s B_{1s}(y^c, y^d) \right)^2 \\
&\quad + o\left(\sum_{s=1}^{r} \left(\frac{\lambda_s}{n} + \lambda_s^2 \right) + \sum_{s=1}^{q} \left(\frac{h_s}{n} + h_s^4 \right) \right)
\end{aligned}$$

Proof of Theorem 2.3.1

We can rewrite the cross-validation function as follows:

$$
\begin{aligned}
CV(\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\int\{I_i - \hat{F}_{-i}\}^2 dy \\
&= \frac{1}{n}\sum_{i=1}^{n}\int\{I_i - F + F - \hat{F}_{-i}\}^2 dy \\
&= \frac{1}{n}\sum_{i=1}^{n}\int\left\{(I_i - F)^2 + (F - \hat{F}_{-i})^2 + 2(I_i - F)(F - \hat{F}_{-i})\right\} dy \\
&= \int\left\{\frac{1}{n}\sum_{i=1}^{n}(I_i - F)^2 + \frac{1}{n}\sum_{i=1}^{n}(F - \hat{F}_{-i})^2 + \frac{2}{n}\sum_{i=1}^{n}(I_i - F)(F - \hat{F}_{-i})\right\} dy \\
&= \int\Bigg[\frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i}^{n}(G_\gamma(y, y_j) - F(y))(G_\gamma(y, y_l) - F(y)) \\
&\quad -\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(G_\gamma(y, y_j) - F(y))(I_i - F) + \frac{1}{n}\sum_{i=1}^{n}(I_i - F)^2\Bigg] dy \\
&\equiv \int\left(S_{1n} - S_{2n} + \frac{1}{n}\sum_{i=1}^{n}(I_i - F)^2\right) dy
\end{aligned}
$$

where

$$
\begin{aligned}
S_{1n} &= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i}^{n}(G_\gamma(y, y_j) - F(y))(G_\gamma(y, y_l) - F(y)) \\
S_{2n} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(G_\gamma(y, y_j) - F(y))(I_i - F)
\end{aligned}
$$

Since the third item is not related to bandwidth γ, we analyze the asymptotic properties of S_{1n} and S_{2n} separately.

$$
\begin{aligned}
S_{1n} &= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i}^{n}(G_\gamma(y,y_j)-F(y))(G_\gamma(y,y_l)-F(y)) \\
&= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(G_\gamma(y,y_j)-F(y))^2 \\
&\quad +\frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq j\neq i}^{n}(G_\gamma(y,y_j)-F(y))(G_\gamma(y,y_l)-F(y)) \\
&\equiv S_{1n,1}+S_{1n,2}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
S_{1n,1} &= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(G_\gamma(y,y_j)-F(y))^2 \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}(G_\gamma(y,y_i)-F(y))^2
\end{aligned}
$$

By law of large number (LLN),

$$
\frac{1}{n}\sum_{i=1}^{n}(G_\gamma(y,y_i)-F(y))^2 \xrightarrow{p} E\left\{(G_\gamma(y,y_i)-F(y))^2\right\}
$$

Therefore, we have

$$
\begin{aligned}
S_{1n,1} &= \frac{1}{n-1}\left[E\left\{(G_\gamma(y,y_i)-F(y))^2\right\}+o_p(1)\right] \\
&= \frac{1}{n-1}\left[\mathrm{Var}\big(G_\gamma(y,y_i)\big)+\left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2+o_p(1)\right] \\
&= \frac{1}{n}\left[\mathrm{Var}\big(G_\gamma(y,y_i)\big)+\left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2\right]+o_p\left(n^{-1}\right) \qquad \text{(A.6)}
\end{aligned}
$$

Similarly, for $S_{1n,2}$, we have

$$\begin{aligned} S_{1n,2} &= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq j\neq i}^{n}(G_\gamma(y,y_j)-F(y))(G_\gamma(y,y_l)-F(y)) \\ &= \frac{n-2}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(G_\gamma(y,y_i)-F(y))(G_\gamma(y,y_j)-F(y)) \end{aligned}$$

Denote $U_1 = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n} H_n(y_i,y_j)$. Then U_1 is a second order U-statistic with the kernel $H_n(y_i,y_j) = (G_\gamma(y,y_i)-F(y))(G_\gamma(y,y_j)-F(y))$. Define $H_{n,i} \equiv E\{H_n(y_i,y_j)|y_i\}$ and $\theta = E\{H_n(y_i,y_j)\}$. Then, by U-statistic H-decomposition, we have

$$\begin{aligned} U_1 &= \theta + \frac{2}{n}\sum_{i=1}^{n}(H_{n,i}-\theta) + \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}(H_n(y_i,y_j)-H_{n,i}-H_{n,j}+\theta) \\ &= J_{1,0}+J_{1,1}+J_{1,2} \end{aligned} \tag{A.7}$$

For each item in equation (A.7), we have

$$\begin{aligned} \theta &= E\{(G_\gamma(y,y_i)-F(y))(G_\gamma(y,y_j)-F(y))\} \\ &= E\{(G_\gamma(y,y_i)-F(y))\}E\{(G_\gamma(y,y_j)-F(y))\} \\ &= \left[E\{(G_\gamma(y,y_i)-F(y))\}\right]^2 \\ &= \left[\text{bias}\left(\hat{F}(y)\right)\right]^2 \\ H_{n,i} &= E\{(G_\gamma(y,y_i)-F(y))(G_\gamma(y,y_j)-F(y))|y_i\} \\ &= (G_\gamma(y,y_i)-F(y))E\{(G_\gamma(y,y_j)-F(y))|y_i\} \\ &= (G_\gamma(y,y_i)-F(y))\text{bias}\left(\hat{F}(y)\right) \\ H_{n,j} &= (G_\gamma(y,y_j)-F(y))\text{bias}\left(\hat{F}(y)\right) \end{aligned}$$

Thus,

$$J_{1,0} = \theta = \left[\text{bias}\left(\hat{F}(y)\right)\right]^2$$

For $J_{1,1}$, we have

$$\begin{aligned}
EJ_{1,1} &= E(H_{n,i}) = 0 \\
EJ_{1,1}^2 &= \text{Var}(J_{1,1}) \\
&= \frac{4}{n}\text{Var}(H_{n,i} - \theta) \\
&= \frac{4}{n}E(H_{n,i} - \theta)^2 \\
&= \frac{4}{n}\left[E(G_\gamma(y, y_i) - F(y))^2 - \left[\text{bias}\left(\hat{F}(y)\right)\right]^2\right]\left[\text{bias}\left(\hat{F}(y)\right)\right]^2 \\
&= \frac{4}{n}\text{Var}\big(G_\gamma(y, y_i)\big)\left[\text{bias}\left(\hat{F}(y)\right)\right]^2
\end{aligned}$$

Given that $\text{Var}\big(G_\gamma(y, y_i)\big) < C$ for some $C \in R$, we have $J_{1,1} = O_p\left(n^{-\frac{1}{2}}\right)J_{1,0} = o_p(J_{1,0})$. Similarly, we can have $J_{1,2} = o_p(J_{1,0})$.

So we have the result about U_1 as

$$U_1 = \left[\text{bias}\left(\hat{F}(y)\right)\right]^2(1 + o_p(1)) \tag{A.8}$$

Notice that $S_{1n,2} = \frac{n-2}{n-1}U_1$. We have

$$\begin{aligned}
S_{1n,2} &= \frac{n-2}{n-1}U_1 \\
&= \frac{n-2}{(n-1)}\left[\text{bias}\left(\hat{F}(y)\right)\right]^2(1 + o_p(1)) \\
&= \left[\text{bias}\left(\hat{F}(y)\right)\right]^2(1 + o_p(1))
\end{aligned} \tag{A.9}$$

Combining the results of equations (A.6) and (A.9), we have

$$
\begin{aligned}
S_{1n} &= \left\{ \frac{1}{n} \left[\mathrm{Var}\big(G_\gamma(y, y_j)\big) + \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \right] + \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \right\} (1 + o_p(1)) \\
&= \left\{ \frac{1}{n} \mathrm{Var}\big(G_\gamma(y, y_j)\big) + \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \right\} (1 + o_p(1)) \qquad \text{(A.10)}
\end{aligned}
$$

For S_{2n}, denote $\sum_{j\neq i, m\neq l}^{n} \equiv \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{j\neq i}^{n}\sum_{m\neq l}^{n}$, we have the following results:

$$
\mathrm{E}S_{2n} = \tfrac{1}{n(n-1)} \textstyle\sum_{i=1}^{n}\sum_{j\neq i}^{n} E\big\{(G_\gamma(y, y_j) - F)(I_i - F)\big\} = 0
$$

and

$$
\begin{aligned}
& \mathrm{E}S_{2n}^2 \\
= {} & \frac{1}{n^2(n-1)^2} \mathrm{E} \left\{ \sum_{j\neq i, m\neq l}^{n} (G_\gamma(y, y_j) - F)(I_i - F)(G_\gamma(y, y_m) - F)(I_l - F) \right\} \\
= {} & \frac{1}{n^2(n-1)^2} \mathrm{E} \left\{ \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{m\neq l}^{n} (G_\gamma(y, y_j) - F)(I_i - F)^2(G_\gamma(y, y_m) - F) \right\} \\
& + \frac{1}{n^2(n-1)^2} \mathrm{E} \left\{ \sum_{i=1}^{n}\sum_{l\neq i}^{n} (G_\gamma(y, y_i) - F)(I_i - F)(G_\gamma(y, y_l) - F)(I_l - F) \right\} \\
= {} & \frac{1}{n(n-1)} \left\{ \mathrm{E}(I_i - F)^2 \left[E\left\{(G_\gamma(y, y_j) - F)^2\right\} + \frac{n-2}{2} \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \right] \right\} \\
& + \frac{1}{2n(n-1)} \big[\mathrm{E}\left\{(G_\gamma(y, y_i) - F)(I_i - F)\right\} \big]^2 \\
= {} & \frac{F(1-F)}{n(n-1)} \left\{ \mathrm{Var}\big(G_\gamma(y, y_j)\big) + \frac{n}{2} \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2 \right\} \\
& + \frac{1}{2n(n-1)} \big[\mathrm{E}\left\{(G_\gamma(y, y_i) - F)(I_i - F)\right\} \big]^2
\end{aligned}
$$

Since

$$\mathrm{E}\left\{(G_\gamma(y,y_i)-F)(I_i-F)\right\} \le \frac{1}{2}\mathrm{E}\left\{(G_\gamma(y,y_i)-F)^2+(I_i-F)^2\right\}$$

Therefore, we have

$$S_{2n} = o_p\left(\left\{\frac{1}{n}\mathrm{Var}\big(G_\gamma(y,y_j)\big) + \left[\mathrm{bias}\left(\hat{F}(y)\right)\right]^2\right\} + \frac{1}{n}F^2(1-F)^2\right) \tag{A.11}$$

Combining the results of equations (A.10) and (A.11), we get the results in Theorem 2.3.1.

Proofs of Theorem 2.3.2.

Let $a_s = h_s n^{1/3}$ for $s = 1,\ldots,q$ and $b_s = \lambda_s n^{2/3}$ for $s = 1,\ldots,r$. Then, selecting $\gamma = (h,\lambda)$ to minimize $CV_L(\gamma)$ is equivalent to choosing $(a,b) \equiv (a_1,\ldots,a_q,b_1,\ldots,b_r)$ to minimize

$$\begin{aligned}&\Upsilon(a,b)\\ =&\int\left[-\alpha_0\sum_{s=1}^{q}a_sF_s(y^c,y^d)+\left(\frac{\kappa_2}{2}\sum_{s=1}^{q}a_s^2F_{ss}(y^c,y^d)+\sum_{s=1}^{r}b_sB_{1s}(y^c,y^d)\right)^2\right]dy\end{aligned}$$

Let(a^0,b^0) denote the values of (a,b) that minimize $\Upsilon(a,b)$ subject to the constraint that each of them is non-negative. We require that

Each a_s^0 is positive and b_s^0 is non-negative, all are finite and uniquely defined.

Thus, we get the results that $n^{1/3}\hat{h}_s \xrightarrow{p} a_s^0$ for $s = 1,\ldots,q$ and $n^{2/3}\hat{\lambda}_s \xrightarrow{p} b_s^0$ for $s = 1,\ldots,r$, which implies the conclusions of theorem 2.

Proofs of Theorem 2.3.3.

Denote $\bar{F}(y)$ to be the estimator when we use $\gamma^0 = (h^0, \lambda^0)$ as the bandwidths for our estimation. Then

$$\bar{F}_{\gamma^0}(y) = \frac{1}{n}\sum_{j=1}^{n} G_{\gamma^0}(y, y_j) \tag{A.12}$$

$\hat{D}_\gamma(y) \equiv \Big[\hat{F}_\gamma(y) - F(y) - \big(\frac{\kappa_2}{2}\sum_{s=1}^{q} h_s^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s B_{1s}(y^c, y^d)\big)\Big]$ has been shown to have asymptotic mean zero and asymptotic variance F(y)(1-F(y)) for any fixed bandwidth γ satisfying the condition 3. Thus, by Liapunov's CLT theorem,

$$\bar{D}_{\gamma^0}(y) \xrightarrow{d} N(0, V), \tag{A.13}$$

where $\bar{D}_{\gamma^0}(y) \equiv \big[\bar{F}_{\gamma^0}(y) - F(y) - \big(\frac{\kappa_2}{2}\sum_{s=1}^{q}(h_s^0)^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \lambda_s^0 B_{1s}(y^c, y^d)\big)\big]$ In theorem 2, we have shown that $\frac{\hat{h}_s}{h_s^0} \xrightarrow{p} 1, \;\; for\ s = 1, \dots, q$; and $\frac{\hat{\lambda}_s}{\lambda_s^0} \xrightarrow{p} 1, \;\; for\ s = 1, \dots, r;$. By using stochastic equicontinuity arguments as in Hall et al. (2004), one can show that $\hat{D}_{\hat{\gamma}}(y) - \bar{D}_{\gamma^0}(y) = o_p(\sqrt{n})$. Thus,

$$\hat{D}_{\hat{\gamma}}(y) \xrightarrow{d} N(0, V), \tag{A.14}$$

where $\hat{D}_{\hat{\gamma}}(y) \equiv \Big[\bar{F}_{\hat{\gamma}}(y) - F(y) - \Big(\frac{\kappa_2}{2}\sum_{s=1}^{q}(\hat{h}_s)^2 F_{ss}(y^c, y^d) + \sum_{s=1}^{r} \hat{\lambda}_s B_{1s}(y^c, y^d)\Big)\Big]$.

Therefore, we have

$$\big[\bar{F}_{\hat{\gamma}}(y) - F(y)\big] \xrightarrow{d} N(0, V). \tag{A.15}$$

APPENDIX B

PROOFS OF THEOREMS IN SECTION 3

Proof of Theorem 3.3.1

We will first prove a lemma (Lemma 1) below which will be used to Theorem 3.3.1.

Lemma 1 Under Assumptions A-C, we have for some $M_1 < \infty$, and for all N and T,

(i) $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 \leq M_1$;
(ii) $E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|(Nr)^{-\frac{1}{2}}e_t'\Lambda^0\right\|^2\right) = E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|(Nr)^{-\frac{1}{2}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right) \leq M_1$;
(iii) $E\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right) \leq M_1$;
(iv) $E\left\|(NTr)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\right\| \leq M_1$

Proof :

(i) Same as Lemma 1(i) in Bai and Ng (2002).

(ii)

$$
\begin{aligned}
&E\left(\left\|(Nr)^{-\frac{1}{2}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right)\\
=&\frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{N}E(e_{it}e_{jt})E(\lambda_i^{0\prime}\lambda_j^0)\\
=&\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij,t}E\left(\frac{\lambda_i^{0\prime}\lambda_j^0}{r}\right)\\
\leq& CM
\end{aligned}
$$

by Assumptions B and C(3).

(iii) Same as Lemma 1(iii) in Bai and Ng (2002).

(iv)

$$
\begin{aligned}
& E\left\|(NTr)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\right\|^2 \\
=& \frac{1}{NTr}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_{it}e_{js})E\left(\lambda_i^{0\prime}\lambda_j^0\right) \\
\leq & C\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \\
\leq & CM
\end{aligned}
$$

by Assumptions B and C(4). ∎

Proof of Theorem 3.3.1:

Recall that $\hat{F}^k = N^{-1}X\tilde{\Lambda}^k$ and $\tilde{\Lambda}^k = T^{-1}X'\tilde{F}^k$. From the normalization $I_k = \tilde{F}^{k\prime}\tilde{F}^k/T$, we also have $(Tk)^{-1}\sum_{t=1}^{T}\|\tilde{F}_t^k\|^2 = 1$. Following Bai and Ng (2002), $H^{k\prime} = (\tilde{F}^{k\prime}F^0/T)(\Lambda^{0\prime}\Lambda^0/N)$, we have

$$
\hat{F}_t^k - H^{k\prime}F_t^0 = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st}
$$

where $\zeta_{st} = e_s'e_t/N - \gamma_N(s,t)$, $\eta_{st} = F_s^{0\prime}\Lambda^{0\prime}e_t/N$, and $\xi_{st} = F_t^{0\prime}\Lambda^{0\prime}e_s/N = \eta_{ts}$.

Because $(x+y+z+u)^2 \leq 4(x^2+y^2+z^2+u^2)$, $\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 \leq 4(a_t + b_t + c_t + d_t)$, where $a_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t)\right\|^2$, $b_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st}\right\|^2$, $c_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st}\right\|^2$ and $d_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st}\right\|^2$. It follows that $(1/T)\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 \leq (4/T)\sum_{t=1}^{T}(a_t+b_t+c_t+d_t)$.

We have $\|\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t)\|^2 \leq \left(\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\cdot\left(\sum_{s=1}^{T}\gamma_N(s,t)^2\right)$ by Cauchy's inequality. Thus,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} a_t &\leq \frac{k}{T}\left(\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\cdot\frac{1}{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N(s,t)^2\right)\\ &= O_p(k/T)\end{aligned}$$

by Lemma 1(i)and the fact that $(Tk)^{-1}\sum_{t=1}^{T}\|\tilde{F}_t^k\|^2 = 1$ (this follows from $I_k = \tilde{F}^{k\prime}\tilde{F}^k/T$).

For b_t, we have that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} b_t &= \frac{1}{T^3}\sum_{t=1}^{T}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st}\right\|^2\\ &= \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\tilde{F}_s^{k\prime}\tilde{F}_u^k\zeta_{st}\zeta_{ut}\\ &\leq \frac{1}{T}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}(\tilde{F}_s^{k\prime}\tilde{F}_u^k)^2\right)^{1/2}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2}\\ &\leq \frac{k}{T}\left(\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2}\\ &= k\left[\frac{1}{T^4}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2}\\ &= O_p\left(\frac{k}{N}\right),\end{aligned}$$

where as shown in Bai and Ng (2002), the last equality follows from the result that $\left[\frac{1}{T^4}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2} = O_p(N^{-1})$.

From $E(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut})^2 = E(\sum_{t=1}^{T}\sum_{v=1}^{T}\zeta_{st}\zeta_{ut}\zeta_{sv}\zeta_{uv}) \leq T^2\max_{s,t}E|\zeta_{st}|^4$ and

$$E|\zeta_{st}|^4 = \frac{1}{N^2}E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^4 \leq \frac{1}{N^2}M$$

by Assumption C5, we have

$$\frac{1}{T}\sum_{t=1}^{T} b_t \leq O_p(k)\frac{1}{T}\sqrt{\frac{T^2}{N^2}} = O_p\left(\frac{k}{N}\right).$$

For c_t, we have

$$\begin{aligned}
c_t &= \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k \eta_{st}\right\|^2 \\
&= \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k F_s^{0\prime}\Lambda^{0\prime} e_t/N\right\|^2 \\
&\leq \frac{1}{N^2}\|e_t'\Lambda^0\|^2\left(\frac{k}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\left(\frac{r}{Tr}\sum_{s=1}^{T}\|F_s^0\|^2\right) \\
&= \frac{1}{N^2}\|e_t'\Lambda^0\|^2 O_p(kr)
\end{aligned}$$

because $\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2 = 1$ and $\frac{r}{Tr}\sum_{s=1}^{T}\|F_s^0\|^2 = O_p(1)$.

It follows that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} c_t &= O_p(kr)\frac{r}{N}\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 \\
&= O_p\left(\frac{kr^2}{N}\right)
\end{aligned}$$

because $\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 = O_p(1)$ by Lemma 1(ii).

The term $(1/T)\sum_{t=1}^{T} d_t = O_p\left(\frac{kr^2}{N}\right)$ can be proved similarly. Combining the above results, we have shown that

$$\begin{aligned}
(1/T)\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 &\leq (4/T)\sum_{t=1}^{T}(a_t + b_t + c_t + d_t) \\
&= O_p\left(\frac{r^2 k}{N}\right) + O_p\left(\frac{k}{T}\right).
\end{aligned}$$

■

Proof of Theorem 3.3.2

Lemma 2 Let $D_k = \frac{\hat{F}^{k\prime}\hat{F}^k}{T}$ and $D_0 = \frac{H^{k\prime}F^{0\prime}F^0H^k}{T}$. When $k \leq r$, we have (i) $\|D_k^{-1}\| = O_p(k)$; (ii) $\|D_k^{-1} - D_0^{-1}\| = O_p\left(\max\left\{\frac{r^2k_{max}^3}{\sqrt{N}}, \frac{rk_{max}^3}{\sqrt{T}}\right\}\right)$.

Proof : Following Bai and Ng (2002), we have

$$\begin{aligned}
D_k - D_0 &= \frac{\hat{F}^{k\prime}\hat{F}^k}{T} - \frac{H^{k\prime}F^{0\prime}F^0H^k}{T} \\
&= \frac{1}{T}\sum_{t=1}^{T}[\hat{F}_t^k\hat{F}_t^{k\prime} - H^{k\prime}F_t^0F_t^{0\prime}H^k] \\
&= \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)(\hat{F}_t^k - H^{k\prime}F_t^0)' + \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)F_t^{0\prime}H^k \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}H^{k\prime}F_t^0(\hat{F}_t^k - H^{k\prime}F_t^0)'.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\|D_k - D_0\| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 + 2\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2\right)^{1/2} \\
=\ & O_p\left(\max\left\{\frac{r^2k}{N}, \frac{k}{T}\right\}\right) + O_p\left(\max\left\{\frac{r\sqrt{k}}{\sqrt{N}}, \frac{\sqrt{k}}{\sqrt{T}}\right\}\right)\cdot O_p\left(r\sqrt{k}\right) \\
=\ & O_p\left(\max\left\{\frac{r^2k}{\sqrt{N}}, \frac{rk}{\sqrt{T}}\right\}\right)
\end{aligned}$$

by Theorem 3.3.1 and the fact that $\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2 = O_p(r^2k_{max})$, which is shown below.

From weakly dependent process of F_t^0, it is easy to show that

$$\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2 - E\left[\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2\right] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{k}\left(\sum_{j=1}^{r}H_{lj}^{k'}F_{tj}^0\right)^2\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{k}\sum_{j=1}^{r}\sum_{i=1}^{r}E\left[H_{lj}^{k'}F_{tj}^0 H_{li}^{k'}F_{ti}^0\right] \\ &= O_p(r^2 k_{max}), \end{aligned}$$

where the last equality uses the fact that $E\left[H_{lj}^{k'}F_{tj}^0 H_{li}^{k'}F_{ti}^0\right]$ is finite for all $t = 1, \ldots, T$, and $l, j, i = 1, \ldots, r$. Thus, we have shown that $\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2 = O_p(r^2 k_{max})$.

Since $\widehat{F}^k = \bar{F}^k(\bar{F}^{k'}\bar{F}^k)^{(1/2)}$ and $\bar{F}^k = X\bar{\Lambda}^k/N$,

$$\begin{aligned} \frac{\widehat{F}^{k'}\widehat{F}^k}{T} &= \frac{1}{T}(\bar{F}^{k'}\bar{F}^k)^{(1/2)}\bar{F}^{k'}\bar{F}^k(\bar{F}^{k'}\bar{F}^k)^{(1/2)} \\ &= \frac{1}{T}(\bar{F}^{k'}\bar{F}^k)^2 \\ &= \frac{1}{TN^4}\left(\bar{\Lambda}^{k'}X'X\bar{\Lambda}^k\right)^2 \\ &= \frac{1}{TN^4}\left(\bar{\Lambda}^{k'}\bar{\Lambda}^k D_{EV}\right)^2 \\ &= \left(\frac{D_{EV}}{N}\right)^2, \end{aligned}$$

where D_{EV} is the $k \times k$ diagonal matrix consisting of $(ev_{(1)}(X'X), \ldots, ev_{(k)}(X'X))$. Bai and Ng (2008) have a similar identity. This implies $k^{-1}\|D_k^{-1}\| = O_p(1)$. Similarly, given $H^k = \frac{\Lambda^{0'}\Lambda^0}{N}\frac{F^{0'}\tilde{F}^k}{T}$, we have that $k^{-1}\|D_0^{-1}\| = O_p(1)$.

From $D_k^{-1} - D_0^{-1} = D_k^{-1}(D_0 - D_k)D_0^{-1}$, we have

$$\begin{aligned}
\|D_k^{-1} - D_0^{-1}\| &= \|D_k^{-1}(D_o - D_k)D_0^{-1}\| \\
&\leq \|D_k^{-1}\| \cdot \|D_0 - D_k\| \cdot \|D_0^{-1}\| \\
&= k^2 \frac{\|D_k^{-1}\|}{k} \cdot \|D_0 - D_k\| \cdot \frac{\|D_0^{-1}\|}{k} \\
&= k^2 \cdot O_p(1) \cdot O_p\left(\max\left\{\frac{kr^2}{\sqrt{N}}, \frac{kr}{\sqrt{T}}\right\}\right) \\
&= O_p\left(\max\left\{\frac{r^2 k_{max}^3}{\sqrt{N}}, \frac{r k_{max}^3}{\sqrt{T}}\right\}\right).
\end{aligned}$$

■

Lemma 3 For $1 \leq k \leq r$, and the H^k defined in Theorem 3.3.1, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{r^4 k_{max}^4}{\sqrt{N}}, \frac{r^3 k_{max}^4}{\sqrt{T}}\right\}\right).$$

Proof :

For the true factor matrix with r factors and H^k defined in Theorem 3.3.1, let $M_{FH}^0 = I - P_{FH}^0$ denote the idempotent matrix spanned by null space of $F^0 H^k$, with $P_{FH^0} = F^0 H^k \left(H^{k\prime} F^{0\prime} F^0 H^k\right)^{-1} H^{k\prime} F^{0\prime}$. Correspondingly, let $M_{\hat{F}}^k = I_T - \hat{F}^k (\hat{F}^{k\prime} \hat{F}^k)^{-1} \hat{F}^{k\prime} = I_T - P_{\hat{F}}^k$. Then

$$\begin{aligned}
V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' M_{\hat{F}}^k \underline{X}_i, \\
V(k, F^0 H^k) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' M_{FH}^0 \underline{X}_i, \\
V(k, \hat{F}^k) - V(k, F^0 H^k) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' (P_{FH}^0 - P_{\hat{F}}^k) \underline{X}_i.
\end{aligned}$$

Following Bai and Ng (2002), let $D_k = \hat{F}^{k\prime}\hat{F}^k/T$ and $D_0 = H^{k\prime}F^{0\prime}F^0H^k/T$. Then

$$
\begin{aligned}
& P^k_{\hat{F}} - P^0_{FH} \\
= \; & \frac{1}{T}\hat{F}^k\left(\frac{\hat{F}^{k\prime}\hat{F}^k}{T}\right)^{-1}\hat{F}^{k\prime} - \frac{1}{T}F^0H^k\left(\frac{H^{k\prime}F^{0\prime}F^0H^k}{T}\right)^{-1}H^{k\prime}F^{0\prime} \\
= \; & \frac{1}{T}[\hat{F}^{k\prime}D_k^{-1}\hat{F}^k - F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}] \\
= \; & \frac{1}{T}\left[(\hat{F}^k - F^0H^k + F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k + F^0H^k)' - F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}\right] \\
= \; & \frac{1}{T}[(\hat{F}^k - F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k)' + (\hat{F}^k - F^0H^k)D_k^{-1}H^{k\prime}F^{0\prime} \\
& + F^0H^kD_k^{-1}(\hat{F}^k - F^0H^k)' - F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}].
\end{aligned}
$$

Thus, $N^{-1}T^{-1}\sum_{i=1}^N \underline{X}_i'(P^k_{\hat{F}} - P^0_{FH})\underline{X}_i = I + II + III + IV$. We consider each term in turn.

$$
\begin{aligned}
I &= \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T(\hat{F}^k_t - H^{k\prime}F^0_t)'D_k^{-1}(\hat{F}^k_s - H^{k\prime}F^0_s)X_{it}X_{is} \\
&\leq \left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T(\hat{F}^k_t - H^{k\prime}F^0_t)'D_k^{-1}(\hat{F}^k_s - H^{k\prime}F^0_s)\right)^{1/2} \\
&\quad \times\left(\frac{1}{T^2}\sum_{t=1}\sum_{s=1}\left(\frac{1}{N}\sum_{i=1}^N X_{it}X_{is}\right)^2\right)^{1/2} \\
&\leq \left(\frac{1}{T}\sum_{t=1}^T\|\hat{F}^k_t - H^{k\prime}F^0_t\|^2\right)\cdot\|D_k^{-1}\|\cdot O_P(1) \\
&= O_p\left(\max\left\{\frac{kr^2}{N},\frac{k}{T}\right\}\right)\cdot k\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^2r^2}{N},\frac{k^2}{T}\right\}\right).
\end{aligned}
$$

by Theorem 3.3.1, Lemma 1(iii) and Lemma 2(i).

$$
\begin{aligned}
II &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t^k - H^{k'}F_t^0)'D_k^{-1}H^{k'}F_s^0 X_{it}X_{is} \\
&\leq \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2\|H^{k'}F_s^0\|^2\cdot\|D_k^{-1}\|^2\right)^{1/2} \\
&\quad \times\left(\frac{1}{T^2}\sum_{t=1}\sum_{s=1}\left(\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right)^{1/2} \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2\right)^{1/2}\cdot\|D_k^{-1}\|\cdot\left(\frac{kr^2}{Tkr^2}\sum_{s=1}^{T}\|H^{k'}F_s^0\|^2\right)^{1/2} \\
&= O_p\left(\max\left\{\left(\frac{kr^2}{N}\right)^{1/2},\left(\frac{k}{T}\right)^{1/2}\right\}\right)\cdot k\cdot k^{1/2}r\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^2r^2}{\sqrt{N}},\frac{k^2r}{\sqrt{T}}\right\}\right).
\end{aligned}
$$

It can be verified that III is also $O_p\left(\max\left\{\frac{k^2r^2}{\sqrt{N}},\frac{k^2r}{\sqrt{T}}\right\}\right)$.

$$
\begin{aligned}
IV &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t^{0'}H^k(D_k^{-1}-D_0^{-1})H^{k'}F_s^0 X_{it}X_{is} \\
&\leq \|D_k^{-1}-D_0^{-1}\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|\cdot|X_{it}|\right)^2 \\
&\leq \|D_k^{-1}-D_0^{-1}\|\frac{kr^2}{N}\sum_{i=1}^{N}\left(\frac{1}{T\sqrt{kr}}\sum_{t=1}^{T}\|H^{k'}F_t^0\|\right)^2 \\
&= \|D_k^{-1}-D_0^{-1}\|\cdot kr^2\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^4r^4}{\sqrt{N}},\frac{k^4r^3}{\sqrt{T}}\right\}\right).
\end{aligned}
$$

where $\|D_k^{-1}-D_0^{-1}\| = O_p\left(\max\left\{\frac{k^3r^2}{\sqrt{N}},\frac{k^3r}{\sqrt{T}}\right\}\right)$ by Lemma 2(ii).

Thus, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{r^4 k_{max}^4}{\sqrt{N}}, \frac{r^3 k_{max}^4}{\sqrt{T}}\right\}\right).$$

■

Lemma 4 For the matrix H^k defined in Theorem 3.3.1, and for each k with $k < r = r_{N,T} \to \infty$, there exists a positive constant C such that

$$\text{plim} \inf_{N,T\to\infty} \inf_k [V(k, F^0 H^k) - V(r, F^0)] \geq C > 0.$$

Proof :

$$\begin{aligned}
& V(k, F^0 H^k) - V(r, F^0) \\
= \; & \frac{1}{NT}\sum_{i=1}^N \underline{X}_i' M_{FH}^0 \underline{X}_i - \frac{1}{NT}\sum_{i=1}^N \underline{X}_i' M_F^0 \underline{X}_i \\
= \; & \frac{1}{NT}\sum_{i=1}^N (F^0\lambda_i^0 + \underline{e}_i)' M_{FH}^0 (F^0\lambda_i^0 + \underline{e}_i) - \frac{1}{NT}\sum_{i=1}^N \underline{e}_i' M_F^0 \underline{e}_i \\
= \; & \frac{1}{NT}\sum_{i=1}^N \lambda_i^{0\prime} F^{0\prime} M_{FH}^0 F^0 \lambda_i^0 + \frac{2}{NT}\sum_{i=1}^N \underline{e}_i' M_{FH}^0 F^0 \lambda_i^0 \\
& + \frac{1}{NT}\sum_{i=1}^N \underline{e}_i' (P_F^0 - P_{FH}^0) \underline{e}_i \\
= \; & I + II + III.
\end{aligned}$$

Notice that $P_F^0 - P_{FH}^0 \geq 0$, thus $III \geq 0$. For the first term,

$$\begin{aligned}
I &= \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{FH}^0F^0\lambda_i^0 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\left(M_{FH}^0F^0\lambda_i^0\right)^{\prime}M_{FH}^0F^0\lambda_i^0 \\
&\geq C > 0
\end{aligned}$$

because $k < r$ and $M_{FH}^0F^0\lambda_i^0 \neq 0$.

Next,

$$II = \frac{2}{NT}\sum_{i=1}^{N}\underline{e}_i^{\prime}F^0\lambda_i^0 - \frac{2}{NT}\sum_{i=1}^{N}\underline{e}_i^{\prime}P_{FH}^0F^0\lambda_i^0.$$

Consider the first term

$$\begin{aligned}
&\left|\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i^{\prime}F^0\lambda_i^0\right| \\
=& \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{e}_{it}F_t^{0\prime}\lambda_i^0\right| \\
\leq& \left(\frac{1}{Tr}\sum_{t=1}^{T}\|F_t^0\|^2\right)^{1/2}\cdot r^{1/2}\cdot r^{1/2}\frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{Nr}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right)^{1/2} \\
=& O_p\left(\frac{r}{\sqrt{N}}\right),
\end{aligned}$$

where the last equality follows from Lemma 1(ii). The second term is also $o_p(1)$, and hence $II = o_p(1)$. ■

Lemma 5 For $r \leq k \leq k_{max}$, $V(k, \hat{F}^k) - V(r, \hat{F}^r) = O_p\left(\max\left\{\frac{r^4k_{max}^2}{N}, \frac{r^2k_{max}}{T}\right\}\right)$.

Proof :

$$\begin{aligned} |V(k,\hat{F}^k) - V(r,\hat{F}^r)| &\leq |V(k,\hat{F}^k) - V(r,F^0)| + |V(r,F^0) - V(r,\hat{F}^r)| \\ &\leq 2\max_{r\leq k} |V(k,\hat{F}^k) - V(r,F^0)|. \end{aligned}$$

Thus, it is sufficient to prove for each k with $r \leq k \leq k_{max}$,

$$V(k,\hat{F}^k) - V(r,F^0) = O_p\left(\max\left\{\frac{r^4k_{max}^2}{N}, \frac{r^2k_{max}}{\sqrt{T}}\right\}\right).$$

Let H^k be as defined in Theorem 3.3.1, with full row rank. Let the $k\times r$ matrix H^{k+} be the generalized inverse of H^k such that $H^kH^{k+} = I_r$. From $\underline{X}_i = F^0\lambda_i^0 + \underline{e}_i$, we have $\underline{X}_i = F^0H^kH^{k+}\lambda_i^0 + \underline{e}_i$. This implies

$$\begin{aligned} \underline{X}_i &= \hat{F}^kH^{k+}\lambda_i^0 + \underline{e}_i - (\hat{F}^k - F^0H^k)H^{k+}\lambda_i^0 \\ &= \hat{F}^kH^{k+}\lambda_i^0 + \underline{u}_i, \end{aligned}$$

where $\underline{u}_i = \underline{e}_i - (\hat{F}^k - F^0H^k)H^{k+}\lambda_i^0$.

Note that

$$
\begin{aligned}
V(k,\hat{F}^k) &= \frac{1}{NT}\sum_{i=1}^{N}\underline{u}_i' M_{\hat{F}}^k \underline{u}_i,\\
V(r,F^0) &= \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' M_{F}^0 \underline{e}_i,
\end{aligned}
$$

$$
\begin{aligned}
& V(k,\hat{F}^k)\\
= \;& \frac{1}{NT}\sum_{i=1}^{N}\left(\underline{e}_i-(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0\right)' M_{\hat{F}}^k \left(\underline{e}_i-(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0\right),\\
= \;& \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' M_{\hat{F}}^k \underline{e}_i - \frac{2}{NT}\sum_{i=1}^{N}\lambda_i^{0'} H^{k+'}(\hat{F}^k-F^0H^k)' M_{\hat{F}}^k \underline{e}_i\\
& +\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0'} H^{k+'}(\hat{F}^k-F^0H^k)' M_{\hat{F}}^k(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0\\
= \;& a+b+c.
\end{aligned}
$$

Because $I-M_{\hat{F}}^k$ is positive semi-definite, $x'M_{\hat{F}}^k x \le x'x$. Thus

$$
\begin{aligned}
c &\le \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0'} H^{k+'}(\hat{F}^k-F^0H^k)'(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0\\
&\le \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i^0\|^2\|H^{k+}\|^2\right)\\
&= O_p\left(\max\left\{\frac{kr^2}{N},\frac{k}{T}\right\}\right)\cdot kr^2\cdot O_p(1)\\
&= O_p\left(\max\left\{\frac{k^2r^4}{N},\frac{k^2r^2}{T}\right\}\right).
\end{aligned}
$$

by Theorem 3.3.1.

For term b, we use the fact that $|\mathrm{tr}(A)| \le r\|A\|$ for any $r \times r$ matrix A. Thus

$$
\begin{aligned}
b &= \frac{2}{T}\mathrm{tr}\left(H^{k+}(\hat{F}^k - F^0H^k)'M^k_{\hat{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\underline{e}_i\lambda_i^0\right)\right) \\
&\le 2r\|H^{k+}\| \cdot \left\|\frac{\hat{F}^k - F^0H^k}{\sqrt{T}}\right\| \cdot \left\|\frac{1}{\sqrt{T}N}\sum_{i=1}\underline{e}_i\lambda_i^0\right\| \\
&\le 2r\|H^{k+}\| \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}^k - F^0H^k\|^2\right)^{1/2} \\
&\quad \times \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}\underline{e}_i\lambda_i^0\right\|^2\right)^{1/2} \\
&= 2r \cdot (kr)^{1/2} \cdot O_p\left(\max\left\{\frac{\sqrt{k}r}{\sqrt{N}}, \frac{\sqrt{k}}{\sqrt{T}}\right\}\right) \cdot r^{1/2} \cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{r^3k}{N}, \frac{r^2k}{T}\right\}\right)
\end{aligned}
$$

by Theorem 3.3.1 and Lemma 1(ii). Therefore,

$$
V(k, \hat{F}^k) = \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'M^k_{\hat{F}}\underline{e}_i + O_p\left(\max\left\{\frac{k^2r^4}{N}, \frac{k^2r^2}{T}\right\}\right).
$$

Thus we have

$$
\begin{aligned}
& V(k, \hat{F}^k) - V(r, F^0) \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'P^0_F\underline{e}_i - \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'P^k_{\hat{F}}\underline{e}_i + O_p\left(\max\left\{\frac{k^2r^4}{N}, \frac{k^2r^2}{T}\right\}\right).
\end{aligned}
$$

Note that

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i &\leq \left\| \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \right\| \cdot \frac{1}{NT^2}\sum_{i=1}^{N}\underline{e}_i' F^0 F^{0\prime}\underline{e}_i \\
&= \left\| \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \right\| \cdot \frac{1}{NT}\sum_{i=1}^{N}\left\| \frac{1}{\sqrt{Tr}}\sum_{t=1}^{T} F_t^0 \underline{e}_{it} \right\|^2 \cdot r \\
&= r \cdot O_p(1) \cdot \frac{1}{T} \cdot r \cdot O_p(1) \\
&= O_p\left(\frac{r^2}{T}\right) \leq O_p\left(\max\left\{\frac{k^2 r^4}{N}, \frac{k^2 r^2}{T}\right\}\right).
\end{aligned}$$

$\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i$ is bounded by the sum of the first k largest eigenvalues of the matrix $A_{NT} = \frac{1}{NT}e'e$, where $e = (e_{ti}), T \times N$. Let $\rho(A)$ denote the largest eigenvalue of a matrix A. Under Assumption C(6), as Bai and Ng (2005) shows, $\rho(A_{NT}) = O_p(C_{NT}^{-2})$, where $C_{NT}^2 = \min(N, T)$. Thus,

$$\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i = O_p\left(\max\left\{\frac{k}{N}, \frac{k}{T}\right\}\right) \leq O_p\left(\max\left\{\frac{k^2 r^4}{N}, \frac{k^2 r^2}{T}\right\}\right).$$

In summary,

$$V(k, \hat{F}^k) - V(r, F^0) = O_p\left(\max\left\{\frac{r^4 k_{max}^2}{N}, \frac{r^2 k_{max}}{T}\right\}\right).$$

■

Proof of Theorem 3.3.2

Proof :

We shall prove that $\lim_{N,T\to\infty} P(PC(k) < PC(r)) = 0$ for all $k \neq r$. Since

$$PC(k) - PC(r) = V(k, \hat{F}^k) - V(r, \hat{F}^r) - (r-k)g(N, T),$$

it is sufficient to prove $P[V(k,\hat{F}^k)-V(r,\hat{F}^r)<(r-k)g(N,T)]\to 0$ as $N,T,k,r\to\infty$.

Consider $k<r$. We have the identity:

$$\begin{aligned}V(k,\hat{F}^k)-V(r,\hat{F}^r) &= [V(k,\hat{F}^k)-V(k,F^0H^k)]+[V(k,F^0H^k)-V(r,F^0H^r)]\\ &\quad+[V(r,F^0H^r)-V(r,\hat{F}^r)].\end{aligned}$$

Lemma 2 implies that the first and the third terms are both $O_p\left(\max\left\{\frac{k_{max}^8}{\sqrt{N}},\frac{k_{max}^7}{\sqrt{T}}\right\}\right)$. Next, we consider the second item. Because F^0H^r and F^0 span the same column space, $V(r,F^0H^r)=V(r,F^0)$. Thus the second item can be rewritten as $V(k,F^0H^k)-V(r,F^0)$, which has a positive limit by Lemma 3. Hence $P[PC(k)<PC(r)]\to 0$ if $(r-k)g(N,T)\to 0$ as $N,T,k,r\to\infty$.

Next, for $k\geq r$,

$$P[PC(k)-PC(r)<0]=P[V(r,\hat{F}^r)-V(k,\hat{F}^k)>(k-r)g(N,T)].$$

By Lemma 4, $V(r,\hat{F}^r)-V(k,\hat{F}^k)=O_p\left(\max\left\{\frac{k_{max}^6}{N},\frac{k_{max}^4}{T}\right\}\right)$. According to our setting, $(k-r)g(N,T)$ converges to zero at a slower rate than $O_p\left(\max\left\{\frac{k_{max}^6}{N},\frac{k_{max}^4}{T}\right\}\right)$. Thus, for $k>r$, $P[PC(k)<PC(r)]\to 0$ as $N,T,k,r\to\infty$. ■